AD-A075 578 STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH F/G 12/1

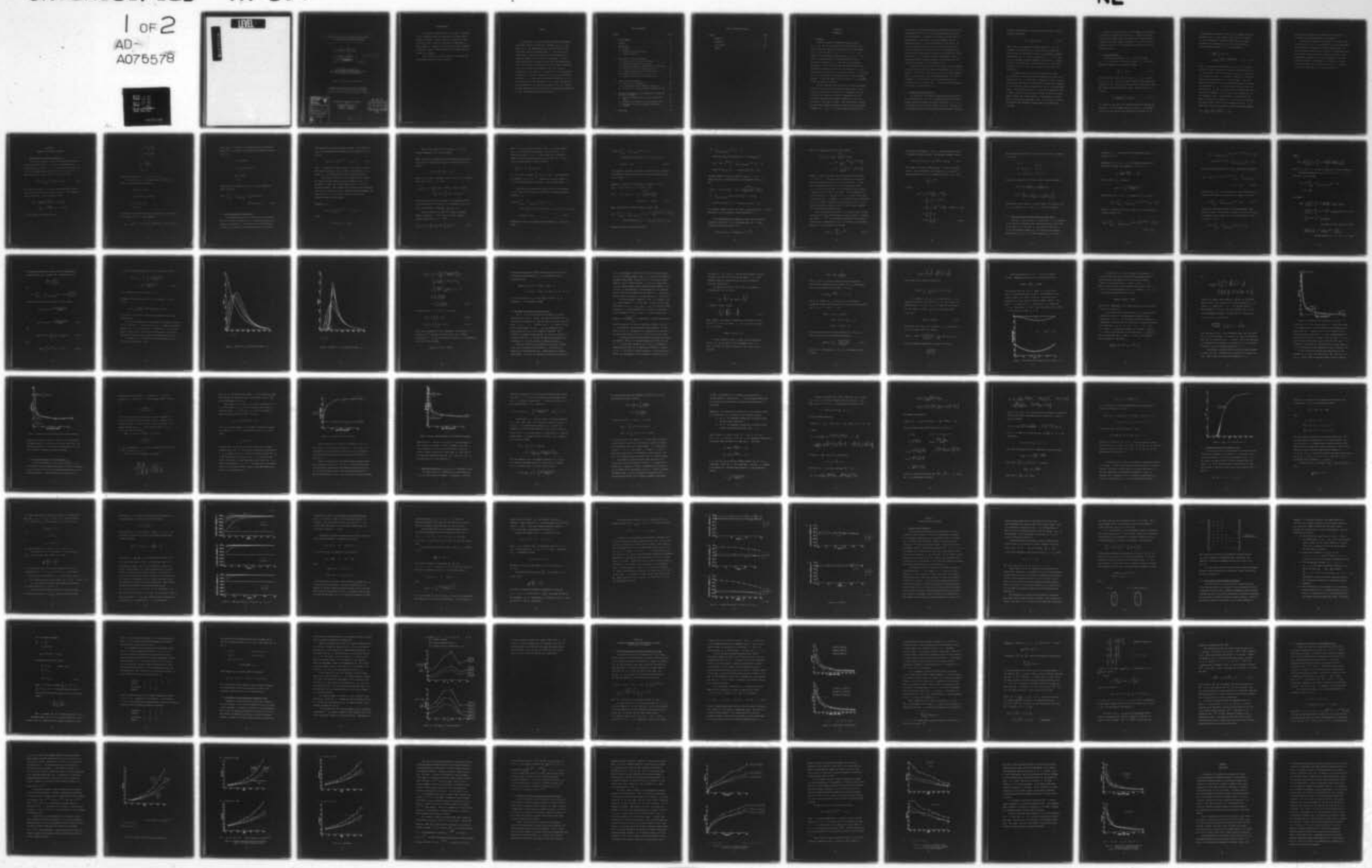
PROPERTIES OF ISOTONIC ESTIMATORS OF MEAN LIFETIME IN A SIMPLE --ETC(U)

AUG 79 T P MCWILLIAMS N00014-75-C-0561

UNCLASSIFIED TR-194 NL

1 OF 2

AD-A075578



LEVEL II

PROPERTIES OF ISOTONIC ESTIMATORS OF MEAN LIFETIME IN A SIMPLE PROTOTYPE DEVELOPMENT MODEL

BY

THOMAS P. McWILLIAMS

TECHNICAL REPORT NO. 194
AUGUST 16, 1979

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-79-C-0561 (NR-047-200)
WITH THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

6

PROPERTIES OF ISOTONIC ESTIMATORS OF MEAN LIFETIME IN A SIMPLE PROTOTYPE DEVELOPMENT MODEL,

BY

10 THOMAS P. McWILLIAMS

9 TECHNICAL REPORT NO. 194 14 TR-194

11 16 AUG[redacted]979 12 113

15

SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-75-C-0561 (NR-047-200)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

402 766

mt

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Gerald J. Lieberman, for his unflagging support and enthusiasm; his patience during periods of scholastic adversity made this dissertation possible. I would also like to thank Dr. Olkin and Dr. Singpurwalla for their suggestions, comments, and service on the reading committee. In addition, thanks to Audrey Stevenin for an excellent and very patient job of typing.

A special thanks to my wife, Connie, for her support and understanding during our years at Stanford.

ABSTRACT

Consideration of a simple prototype development/reliability growth model leads to observation of variables $X_1, X_2, \ldots, X_k$; where the X_i are independent and are assumed to have the same distribution type with individual means λ_i. The problem is to estimate the λ_i, which are assumed to be non-decreasing due to design improvements in the device under development. For many common distributions, estimation via the restricted maximum likelihood principle leads to estimates which are the isotonic regression of the X_i with appropriate weights. This paper finds the distribution of $\hat{\lambda}_k$, the mle of λ_k, when the X_i are exponentially distributed. $\hat{\lambda}_k$ is then compared with $\bar{X}$ and X_k, two competing estimates of λ_k. Gamma, normal, and Weibull distributed X_i are also considered, and the usefulness of $\hat{\lambda}_k$ is seen to depend on the relative magnitude of the variances and means of the X_i. Results are also examined when the restricted mles are generated assuming the λ_i are a non-decreasing and concave function of i.

TABLE OF CONTENTS

TABLE OF CONTENTS Continued

CHAPTER 1

INTRODUCTION

1.1. The Model

Consider a sequence of independent random variables $X_1, \ldots, X_k$. The X_i are assumed to have distributions which belong to the same family, with the possibility of unequal means. Let λ_i denote the mean of X_i, $i = 1, \ldots, k$. The λ_i are unknown, but we assume the ordering relation $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$ holds. The problem treated in this dissertation is that of efficiently estimating the λ_i under various distributional assumptions. A look at the practical considerations from which the model was developed indicates that the estimation of λ_k, the final mean in the sequence, is of particular interest. The case where the X_i are exponentially distributed will be the most thoroughly examined. We will also examine the cases where the X_i have gamma, normal, and Weibull distributions.

The model described above arose from the following situation. Consider a prototype development/testing procedure where a device prototype is tested and its lifetime X_1 is observed. When the prototype fails, the cause of failure is determined and a new prototype is designed in an attempt to eliminate or reduce the probability of this type of failure. This new prototype is tested and the process continued, yielding a sequence of observed lifetime variables $X_1, \ldots, X_k$. If the basic design and construction of

the prototypes is unchanged throughout the development process, then the assumption that the X_i have the same distribution type would not be unreasonable. In addition, if we choose to believe that the design changes at each stage are indeed improvements, then the assumption that the λ_k are non-decreasing is a natural one. The focus of interest on the final mean, λ_k, is due to the fact that this value represents the current state of the development process. A major portion of this paper is devoted to the evaluation of various estimates of λ_k in certain situations, and some work is done on testing hypotheses to determine whether or not a desired mean lifetime has been achieved in the development process.

Note that nothing has been said about the nature of the change in mean lifetime at each stage, beyond the assumption that it is non-negative. For the bulk of the paper this will be the only assumption made - no particular functional relationship among the λ_i will be considered. The only further restricting assumption will be in consideration of the case where the λ_i are not only non-decreasing but are also a concave function of i.

1.2. Maximum Likelihood Estimation

One approach to the parameter estimation problem described in the preceding section is the use of the maximum likelihood principle. Maximum likelihood estimates are given by the values $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ which maximize the joint density function of the X_i. Since the X_i are independent and we have assumed that their distribution functions

belong to the same family, we obtain for the joint density function the simple expression:

$$L(\underline{\nu}) = \prod_{i=1}^{k} f_X(x_i; \underline{\nu}_i) \tag{1.2.1}$$

where $f_X(x_i; \underline{\nu}_i)$ represents the density function of X_i. The assumption of non-decreasing λ_i should naturally be taken into consideration, and consequently we arrive at the problem of maximizing expression (1.2.1) under the assumption $\hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_k$. Note that $\underline{\nu}_i$ represents the entire parameter vector, while λ_i denotes the mean of the distribution. Thus, the value of λ_i may itself be an element of $\underline{\nu}_i$, or it may be a function of one or more of the elements of $\underline{\nu}_i$.

For many densities, the restricted maximization problem described above is difficult and time-consuming, and no closed form expression for the $\hat{\lambda}_i$ can be found. Fortunately, however, there are many important cases where this problem has been shown to be equivalent to that of finding an isotonic regression of the X_i with appropriate weights. This equivalence holds for a fairly broad exponential class of density functions, which includes the normal (with known variance), gamma, binomial, and Poisson densities. In these cases, the values which give the isotonic regression of the X_i are equal to the values which maximize the likelihood function under the non-decreasing constraint.

Section 2.2 includes a proof of the aforementioned equivalence in the case of exponentially distributed X_i. Similar results for the other parent distributions listed above can be found in Chapter 2 of Barlow, Bartholomew, Bremner, and Brunk [1]. Isotonic regression is defined in Section 1.3, and computational algorithms are given in Sections 1.3 and 3.2.

1.3. Isotonic Regression

The isotonic regression of $X_1, \ldots, X_k$ with weights $w_1, \ldots, w_k$ is defined to be the set of values $X_1^*, \ldots, X_k^*$ which minimize the weighted sum of squared deviations

$$\sum_{i=1}^{k} (Y_i - X_i)^2 w_i \tag{1.3.1}$$

where the minimum is taken over all values $Y_1, \ldots, Y_k$ such that $Y_1 \le Y_2 \le \cdots \le Y_k$. There are several relatively simple methods for computing the X_i^*. The primary method used in this paper is the "min-max" formula (Barlow, Bartholomew, Bremner, and Brunk, page 19 [1]), given by

$$X_i^* = \max_{s \le i} \min_{t \ge i} \sum_{r=s}^{t} X_r w_r / \sum_{r=s}^{t} w_r \,. \tag{1.3.2}$$

This formula is very easy to use, involving only the calculation and comparison of various weighted averages of the data. In subsequent sections, we will consider parent distributions for which the X_i^*

(computed using unit weights) will prove to be maximum likelihood estimates of the λ_i, and consequently the X_i^* will be denoted $\hat{\lambda}_i$ for the remainder of the paper. Note that if we consider estimation of the final mean, λ_k, and if all observations are equally weighted, formula (1.3.2) can be reduced to:

$$\hat{\lambda}_k = \max_{s \le k} \sum_{r=s}^{k} X_r/(k-s+1)$$

$$= \max\left\{X_k, \frac{X_k+X_{k-1}}{2}, \frac{X_k+X_{k-1}+X_{k-2}}{3}, \ldots, \bar{X}\right\} \qquad (1.3.3)$$

where $\bar{X}$ represents the overall mean of the X_i. For evaluation purposes, it would be desirable to obtain the distributions of the $\hat{\lambda}_i$ for a variety of parent distributions $f_X(x_i; \underline{\nu}_i)$. This can be very difficult even when k, the number of stages in the process, is small and the density function f is fairly simple. Due to its computationally easier form, some progress has been made on finding the distribution of $\hat{\lambda}_k$. In a paper by David Williams (1977), the distribution of $\hat{\lambda}_k$ and $\hat{\lambda}_1$ (which also has a computationally easier form than the remaining $\hat{\lambda}_i$ 1/) is found in the case where the X_i are independent $N(0,1)$ random variables. No distributions are found in the case where the means λ_i are not equal, but the equal-mean distribution is used to test the hypothesis that the λ_i are indeed equal versus the alternative of unequal non-decreasing means.

1/ $\hat{\lambda}_1 = \min\left\{X_1, \frac{X_1+X_2}{2}, \frac{X_1+X_2+X_3}{3}, \ldots, \bar{X}\right\}$

The problem of determining point estimates of the individual λ_i in the case of non-decreasing means is not addressed.

This paper will develop a formula which can be used to find the distribution of $\hat{\lambda}_k$ in the case where the X_i are independent and exponentially distributed with possibly unequal means. An exact closed-form distribution will be given in the equal-mean case, and hypothesis testing and asymptotics will be discussed. Other distributional assumptions will also be considered, with results primarily obtained through simulation. In addition, a "concave isotonic" technique will be examined, where the λ_i are assumed to be non-decreasing <u>and</u> to be a concave function of i.

CHAPTER II

PROPERTIES OF ISOTONIC ESTIMATORS

2.1. Distribution of the Isotonic Estimator of λ_k

It was seen in Chapter I that it is of particular interest to evaluate the distribution of $\hat{\lambda}_k$ so that we may determine its effectiveness as an estimator of λ_k. The distribution function can be found by integrating the joint density of the X_i over the appropriate region. Letting $F_k(x) = P(\hat{\lambda}_k \leq x)$, we obtain

$$F_k(x) = \iint_{\{\hat{\lambda}_k \leq x\}} \cdots \iint \prod_{i=1}^{k} f_X(x_i; \underline{\nu}_i)\, dx_i \ . \qquad (2.1.1)$$

Now recall that formula (1.3.3) gives a relatively simple form for $\hat{\lambda}_k$ in the case of equal weights, which will prove to be applicable in many situations. Use this formula to obtain

$$\begin{aligned} F_k(x) &= P\left\{\max\left(X_k, \frac{X_k+X_{k-1}}{2}, \ldots, \bar{X}\right) \leq x\right\} \\ &= P\left\{X_k \leq x, \frac{X_k+X_{k-1}}{2} \leq x, \ldots, \bar{X} \leq x\right\} \ . \end{aligned}$$

A natural change in variables is to let

$$Y_1 = \frac{X_1 + \cdots + X_k}{k}$$

$$Y_2 = \frac{X_2 + \cdots + X_k}{k-1}$$

$$\vdots$$

$$Y_{k-1} = \frac{X_{k-1} + X_k}{2}$$

$$Y_k = X_k ,$$

which gives: $F_k(x) = P\{Y_1 \leq x, Y_2 \leq x, \ldots, Y_k \leq x\}$.

Solving for the X_i in the above system of equations we obtain, for the inverse transformation:

$$X_1 = kY_1 - (k-1)Y_2$$

$$\vdots$$

$$X_i = (k-i+1)Y_i - (k-i)Y_{i+1}$$

$$\vdots$$

$$X_{k-1} = 2Y_{k-1} - Y_k$$

$$X_k = Y_k .$$

The Jacobian of this transformation is easily seen to be $k!$, and the joint density of the Y_i can be expressed:

$$f_{Y_1,\ldots,Y_k}(y_1, \ldots, y_k) = k! \prod_{i=1}^{k} f_X((k-i+1)\, y_i - (k-i)\, y_{i+1};\, \underline{\nu}_i)$$

where $y_{k+1} \equiv 0$. Since the X_i represent prototype lifetimes we require $X_i \geq 0$ for each i, which gives the following constraints on the Y_i:

$$
\begin{aligned}
Y_1 &\geq Y_2(k-1)/k \\
&\vdots \\
Y_i &\geq Y_{i+1}(k-i)/(k-i+1) \\
&\vdots \\
Y_{k-1} &\geq Y_k(1/2) \\
Y_k &\geq 0 \ .
\end{aligned}
$$

Integrating the joint density of the Y_i over the appropriate limits, we obtain:

$$
F_k(x) = \int_0^x \int_{y_k/2}^x \cdots \int_{[(k-1)/k]y_2}^x k! \prod_{i=1}^k f_X((k-i+1)y_i - (k-i)y_{i+1}; \underline{\nu}_i)dy_i \ . \qquad (2.1.2)
$$

2.2 Underlying Exponential Distribution

The exponential distribution is widely used in reliability models (cf. Barlow and Prochan [2] for a historical perspective and examples of applications). It was chosen as the first lifetime distribution to consider in this investigation due to its fairly

wide applicability and simple mathematical form. If we assume that the X_i are exponentially distributed, then the density of X_i is given by

$$f_X(x_i; \underline{\nu}_i) = \frac{1}{\lambda_i} e^{-x_i/\lambda_i} \qquad x_i, \lambda_i > 0 \qquad (2.2.1)$$

where λ_i represents the mean as before. In the steps to follow it will be more convenient to work in terms of the reciprocal of λ_i, denoted θ_i. This relationship will hold throughout the paper.

It was stated in Section 1.2 that in many cases the problem of finding restricted maximum likelihood estimates of the λ_i is equivalent to finding the isotomic regression of the X_i with appropriate weights. This statement will be proven in the exponential case, and the reader is referred to Barlow, Bartholomew, Bremner, and Brunk, Section 2.4 [1] for results concerning a more general exponential family of distributions.

Theorem 2.1: Let

$$L(\underline{\lambda}; \underline{x}) = \prod_{i=1}^{k} \frac{1}{\lambda_i} e^{-x_i/\lambda_i} \qquad \lambda_i, x_i > 0$$

and let

$$\Lambda = \{\underline{\lambda} | \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k\} .$$

Then $L(\underline{\hat{\lambda}}; \underline{x}) = \max_{\Lambda} L(\underline{\lambda}; \underline{x})$ if and only if $\underline{\hat{\lambda}}$ is the isotonic regression of $\underline{x}$ with unit weights.

Proof: First note that maximizing the likelihood function $L(\underline{\lambda}; \underline{x})$ is equivalent to minimizing the negative log likelihood function, given by

$$\ell(\underline{\lambda}; \underline{x}) = \sum_{i=1}^{k} (\log \lambda_i + x_i/\lambda_i) .$$

Now let $\Phi(u) = -\log u$ and define $\Delta\Phi(u,v) = \Phi(u) - \Phi(v) - (u-v)\phi(v)$ where $\phi(u) = \Phi'(u) = -1/u$. Then

$$\sum_{i=1}^{k} \Delta_\phi(x_i, \lambda_i) = \sum_{i=1}^{k} \left(-\log x_i + \log \lambda_i - (x_i-\lambda_i)\left(-\frac{1}{\lambda_i}\right)\right)$$

$$= -\sum_{i=1}^{k} \log x_i + \sum_{i=1}^{k} (\log \lambda_i + x_i/\lambda_i) - k .$$

This expression differs from the negative log likelihood function only in terms which do not include $\underline{\lambda}$, so minimizing $\ell(\underline{\lambda}; \underline{x})$ is seen to be equivalent to minimizing $\sum_{i=1}^{k} \Delta_\phi(x_i, \lambda_i)$.

Now apply theorem 1.10, page 41 of Barlow, Bartholomew, Bremner, and Brunk [1] . This theorem, rewritten using our notation, states that if $\underline{\lambda} \in \Lambda$ then

$$\sum_{i=1}^{k} \Delta_\phi(x_i, \lambda_i) \geq \sum_{i=1}^{k} \Delta_\phi(x_i, x_i^*) + \sum_{i=1}^{k} \Delta_\phi(x_i^*, \lambda_i) \qquad (2.2.2)$$

where $\underline{x}^*$ is the isotonic regression of the x_i (with unit weights) and Φ is an arbitrary convex function. Note that since Φ is assumed to be convex, $\Delta_\Phi(u, v)$ is easily seen to be non-negative. Using this fact and applying expression (2.2.2) with $\Phi(u) = -\log u$ (which is indeed convex) we obtain the result:

$$\sum_{i=1}^{k} \Delta_\Phi(x_i, \lambda_i) \geq \sum_{i=1}^{k} \Delta_\Phi(x_i, x_i^*), \qquad \forall\, \underline{\lambda} \in \Lambda .$$

So, $\underline{x}^*$ is seen to minimize $\sum_{i=1}^{k} \Delta_\Phi(x_i, \lambda_i)$ over Λ and consequently $\underline{x}^*$ minimizes $\ell(\underline{\lambda}, \underline{x})$ over Λ. Therefore the isotonic regression $\underline{x}^*$ and the maximum likelihood estimate $\hat{\underline{\lambda}}$ are one and the same. □

Now substitute the exponential density function into expression (2.1.2) to find the distribution of the isotonic or maximum likelihood estimate of λ_k:

$$F_k(x) = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-1)/k]y_2}^x k!\, \theta_1\theta_2 \cdots \theta_k$$

$$\cdot \exp\{-[\theta_1(ky_1 - \{k-1\}\, y_2) + \theta_2(\{k-1\}\, y_2 - \{k-2\}\, y_3) + \cdots + \theta_k y_k]\}$$

$$\cdot\, dy_1\, dy_2 \cdots dy_k \,. \qquad (2.2.3)$$

Combining the terms in the exponent which have common Y_i factors, we obtain:

$$F_k(x) = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-1)/k]y_2}^x k!\theta_1\theta_2 \cdots \theta_k$$

$$\cdot \exp(-[k\theta_1 y_1 + (k-1)(\theta_2-\theta_1)\, y_2 + \cdots + (\theta_k-\theta_{k-1})\, y_k])$$

$$\cdot\, dy_1\, dy_2 \cdots dy_k \;. \qquad (2.2.4)$$

This integral is difficult to evaluate in closed form, but a recursive relationship has been developed which can be used to determine $F_k(x)$ for any value of k.

<u>Theorem 2.2</u>: Denote the integral given in expression (2.2.4) by $a_k(\theta_1, \ldots, \theta_k)$ (consider x to be fixed). Then

$$a_k(\theta_1, \ldots, \theta_k) = a_{k-1}(\theta_2, \ldots, \theta_k) - \frac{e^{-k\theta_1 x}\,\theta_2 \cdots \theta_k}{(\theta_2-\theta_1)\cdots(\theta_k-\theta_1)}$$

$$\cdot\, a_{k-1}(\theta_2-\theta_1, \ldots, \theta_k-\theta_1) \;. \qquad (2.2.5)$$

<u>Proof</u>: Evaluating the innermost integral of $a_k(\underline{\theta})$ gives

$$a_k(\underline{\theta}) = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-2)/(k-1)]y_3}^x (k-1)!\;\theta_2 \cdots \theta_k \left\{ e^{-(k-1)\theta_1 y_2} - e^{-k\theta_1 x} \right\}$$

$$\cdot \exp\{-[(k-1)(\theta_2-\theta_1)\, y_2 + \cdots + (\theta_k-\theta_{k-1})\, y_k]\}\, dy_2 \cdots dy_k \;.$$

Separate this into two integrals to obtain

$$\int_0^x \cdots \int_{[(k-2)/(k-1)]y_3}^x (k-1)!\ \theta_2 \cdots \theta_k$$

$$\cdot \exp\{-[(k-1)\ \theta_2 y_2 + (k-2)(\theta_3-\theta_2)\ y_3 + \cdots + (\theta_k-\theta_{k-1})\ y_k]\}$$

$$\cdot\ dy_2 \cdots dy_k - e^{-k\theta_1 x} \int_0^x \cdots \int_{[(k-2)/(k-1)]y_3}^x (k-1)!\ \theta_2 \cdots \theta_k$$

$$\cdot \exp\{-[(k-1)(\theta_2\ \theta_1)\ y_2 + \cdots + (\theta_k-\theta_{k-1})\ y_k]\}\ dy_2 \cdots dy_k\ .$$

The first integral is seen to be equal to $a_{k-1}(\theta_2, \ldots, \theta_k)$. Substituting this in and multiplying by the appropriate factors, we have

$$a_k(\theta_1, \ldots, \theta_k) = a_{k-1}(\theta_2, \ldots, \theta_k) - \frac{e^{-k\theta_1 x}\ \theta_2 \cdots \theta_k}{(\theta_2-\theta_1)(\theta_3-\theta_1) \cdots\ \theta_k-\theta_1)}\ .$$

$$\int_0^x \cdots \int_{[(k-2)/(k-1)]y_3}^x (k-1)!\ (\theta_2-\theta_1)(\theta_3-\theta_1) \cdots (\theta_k-\theta_1)$$

$$\cdot \exp[-\{(k-1)(\theta_2-\theta_1)\ y_2 + \cdots + (\theta_k-\theta_{k-1})\ y_k\}]\ dy_2 \cdots dy_k\ .$$

The remaining integral is seen to be equal to $a_{k-1}(\theta_2-\theta_1, \theta_3-\theta_1, \ldots, \theta_k-\theta_1)$. Expression (2.2.5) is obtained by direct substitution. □

The relationship given in Theorem 2.2 can be used to find the distribution function of $\hat{\lambda}_k$ for successive values of k . First consider the trivial case, when $k = 1$:

$$a_1(\theta_1) = P\{\hat{\lambda}_1 \le x\} = P\{\max(X_1) \le x\} = 1-e^{-\theta_1 x}\ .$$

When $k = 2$, expression (2.2.5) is used to obtain

$$a_2(\theta_1, \theta_2) = a_1(\theta_2) - \frac{e^{-2\theta_1 x}\theta_2}{\theta_2-\theta_1} a_1(\theta_2-\theta_1)$$

$$= 1 - e^{-\theta_2 x} - \frac{\theta_2}{\theta_2-\theta_1} e^{-2\theta_1 x}\left(1-e^{-(\theta_2-\theta_1)x}\right)$$

$$= 1 - e^{-\theta_2 x} - \frac{\theta_2}{\theta_2-\theta_1}\left[e^{-2\theta_1 x} - e^{-(\theta_1+\theta_2)x}\right].$$

Appendix A gives expressions for the distribution function when $k = 3$ and $k = 4$. This rather tedious process can be continued to find the distribution function of $\hat{\lambda}_k$ for any value of k. Note that the problem of division by zero is encountered above when $\theta_1 = \theta_2$, and this will continue to be a problem for larger values of k whenever any pair of the θ_i are equal. An exact expression for the distribution of $\hat{\lambda}_k$ when <u>all</u> of the θ_i are equal will be found in the next section. In Appendix B it is shown that the distributions for the "in-between" cases where at least one pair of the θ_i are equal can be found by computing $a_k(\underline{\theta})$ via the recursive technique and then taking the appropriate limits.

Once the distribution function of $\hat{\lambda}_k$ has been found, its mean, variance, and any other desired moments are easily computed. Recalling that $F_1(x) = 1 - e^{-\theta_1 x}$ and considering the nature of the recursive relation given in expression (2.2.5), it is seen that the distribution function of $\hat{\lambda}_k$ will always be of the form

$$F_k(x) = 1 - \sum_{i=1}^{2^k-1} \alpha_i e^{-\beta_i x} \qquad (2.2.6)$$

for some real coefficients α_i and β_i. This represents a mixture of exponential density functions. Now consider the moment equation

$$E(X^r) = \int_0^\infty x^r f(x)\,dx = r\int_0^\infty x^{r-1}[1 - F(x)]\,dx \qquad (2.2.7)$$

which holds for a positive random variable X having distribution function $F(x)$ with density $f(x)$ (Feller, Vol. II, page 150 [9]. Applying expression (2.2.7) to a distribution function of the form

$$1 - \sum_{i=1}^{N} \alpha_i e^{-\beta_i x}$$

we obtain

$$E(X^r) = r\int_0^\infty x^{r-1}\left[\sum_{i=1}^{N} \alpha_i e^{-\beta_i x}\right] dx$$

$$= r\sum_{i=1}^{N} \alpha_i \int_0^\infty x^{r-1} e^{-\beta_i x}\,dx$$

$$= r\sum_{i=1}^{N} \alpha_i \int_0^\infty \left(\frac{y}{\beta_i}\right)^{r-1} e^{-y}\,\frac{dy}{\beta_i}\,, \quad \text{(letting } y = \beta_i x)$$

$$= r\sum_{i=1}^{N} \frac{\alpha_i}{\beta_i^r}\Gamma(r)$$

$$= \Gamma(r+1)\sum_{i=1}^{N} \frac{\alpha_i}{\beta_i^r}\,. \qquad (2.2.8)$$

The distribution function $F_2(x)$ is seen to fit the form of expression (2.2.6) with

$$\begin{aligned} \alpha_1 &= 1 & \beta_1 &= \theta_2 \\ \alpha_2 &= \theta_2/(\theta_2 - \theta_1) & \beta_2 &= 2\theta_1 \\ \alpha_3 &= -\theta_2/(\theta_2 - \theta_1) & \beta_3 &= \theta_1 + \theta_2 . \end{aligned}$$

Consequently, expression (2.2.8) can be used to obtain the moments

$$E(\hat{\lambda}_2) = \frac{1}{\theta_2} + \frac{\theta_2}{2\theta_1(\theta_2-\theta_1)} - \frac{\theta_2}{(\theta_1+\theta_2)(\theta_2-\theta_1)}$$

$$E(\hat{\lambda}_2^{\,2}) = 2\left[\frac{1}{\theta_2^2} + \frac{\theta_2}{4\theta_1^2(\theta_2-\theta_1)} - \frac{\theta_2}{(\theta_1+\theta_2)^2(\theta_2-\theta_1)}\right] .$$

Other moments can be computed in a similar manner. The expected value of $\hat{\lambda}_k$ for the cases $k = 3$ and $k = 4$ is given in Appendix A.

2.3. Underlying Exponential Distributions with Equal Means

When the λ_i are all equal to a common value $\lambda = 1/\theta$, expression (2.2.4) simplifies considerably and a closed form solution can be found for any value of k. The equal λ_i case will prove useful for (1) testing hypotheses regarding λ_k; (2) evaluating the asymptotic distribution of $\hat{\lambda}_k$ as $k \to \infty$; and (3) comparing $\hat{\lambda}_k$ with other

estimates of λ_k. These topics will be discussed in detail in Sections 2.4 - 2.7.

Theorem 2.3: Let $X_1, X_2, \ldots, X_k$ be independent exponentially distributed lifetimes with common mean $\lambda = 1/\theta$. Let

$$\hat{\lambda}_k = \max\left(X_k, \frac{X_k+X_{k-1}}{2}, \ldots, \bar{X}\right).$$

Then the CDF of $\hat{\lambda}_k$ is given by

$$F_k(x) = 1 - \sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{i-1} e^{-j\theta x}}{(j-1)!} .$$

Proof: Replace each θ_i by θ in expression (2.2.4) to obtain

$$F_k(x) = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-1)/k]y_2}^x k!\,\theta^k e^{-k\theta y_1} dy_1\, dy_2 \cdots dy_k .$$

Consider x fixed and denote this integral $b_k(\theta)$. Evaluating the innermost integral gives

$$b_k(\theta) = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-2)/(k-1)]y_3}^x (k-1)!\,\theta^{k-1}\left[e^{-(k-1)\theta y_2} - e^{-k\theta x}\right] \cdot dy_2 \cdots dy_k$$

$$= \int_0^x \cdots \int_{[(k-2)/(k-1)]y_2}^x (k-1)!\, \theta^{k-1} e^{-(k-1)\theta y_2}\, dy_2 \cdots dy_k$$

$$- (k-1)!\, \theta^{k-1} e^{-k\theta x} \int_0^x \cdots \int_{[(k-2)/(k-1)]y_3}^x dy_2 \cdots dy_k \,.$$

The first integral given above is $b_{k-1}(\theta)$, giving the relationship

$$b_k(\theta) = b_{k-1}(\theta) - (k-1)!\, \theta^{k-1} e^{-k\theta x} \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-2)/(k-1)]y_3}^x dy_2 \cdots dy_k \,. \tag{2.3.1}$$

Therefore, the $b_k(\theta)$ can be determined if their differences can be evaluated, which requires computation of an integral of the form

$$a_k \equiv \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-2)/(k-1)]y_3}^x dy_2 \cdots dy_k \,. \tag{2.3.2}$$

This integral represents the volume of a multi-dimensional figure bounded by planes. We will derive a formula for a more general class of integrals and then consider (2.3.2) as a special case. Define a function of k and r by

$$c_k^{(r)} = \int_0^1 \int_{y_{k/2}}^1 \cdots \int_{[(k-1)/k]y_2}^1 y_1^r\, dy_1\, dy_2 \cdots dy_k \,.$$

Lemma 1:

$$c_k^{(r)} = \frac{1}{r+1}\left[\frac{k^{k-2}}{\{(k-1)!\}^2} + \sum_{i=1}^{k-1} (-1)^i \frac{(k-i)^{k+r}}{(k-i)!} \frac{(r+1)!}{(r+i+1)!} \frac{1}{k^r k!}\right]$$

where $k \geq 2$ and $r \geq 0$. A proof of Lemma 1 can be found in Appendix B.

Now make the substitution $z_i = y_i/x$ for $i = 2, 3, \ldots, k$ in expression (2.3.2) to obtain

$$a_k = x^{k-1} \int_0^1 \int_{z_{k/2}}^1 \cdots \int_{[(k-2)/(k-1)]z_3}^1 dz_2 \cdots dz_k$$

$$= x^{k-1} c_{k-1}^{(0)} .$$

From Lemma 1,

$$c_{k-1}^{(0)} = \frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{i=1}^{k-2} (-1)^i \frac{(k-1-i)^{k-1}}{(k-1-i)!} \frac{1}{(i+1)!} \frac{1}{(k-1)!}$$

$$= \frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{j=1}^{k-2} (-1)^{k-1-j} \frac{j^{k-1}}{j!} \frac{1}{(k-j)!(k-1)!} \quad \text{(let } j = k-1-i\text{)}$$

$$= \sum_{j=1}^{k-1} (-1)^{k-1-j} \frac{j^{k-1}}{j!} \frac{1}{(k-j)!(k-1)!}$$

(incorporate the first term into the sum)

$$= \frac{(-1)^{k-1}}{k!(k-1)!} \sum_{j=0}^{k} (-1)^{-j} \frac{k!}{j!(k-j)!} j^{k-1} - \frac{k^{k-2}}{[(k-1)!]^2}$$

(add and subtract the $j = 0$ and $j = k$ terms) .

The summation given above is equal to zero by a combinatorial identity (Feller, Vol. I, page 65 [8]). Consequently we have

$$c_{k-1}^{(0)} = \frac{k^{k-2}}{[(k-1)!]^2}$$

and

$$a_k = \int_0^x \int_{y_{k/2}}^x \cdots \int_{[(k-2)/(k-1)]y_3}^x dy_2 \cdots dy_k = \frac{x^{k-1}\, k^{k-2}}{[(k-1)!]^2} \; .$$

Substitute this result into expression (2.3.1) to obtain

$$b_k(\theta) = b_{k-1}(\theta) - \frac{k^{k-2}(\theta x)^{k-1} e^{-k\theta x}}{(k-1)!}$$

or

$$b_k(\theta) - b_{k-1}(\theta) = - \frac{k^{k-2}(\theta x)^{k-1} e^{-k\theta x}}{(k-1)!} \; . \tag{2.3.3}$$

Now

$$b_k(\theta) = b_1(\theta) + \sum_{j=2}^{k} (b_j(\theta) - b_{j-1}(\theta)) \tag{2.3.4}$$

and

$$b_1(\theta) = \int_0^x \theta e^{-\theta y_1}\, dy_1 = 1 - e^{-\theta x} \tag{2.3.5}$$

so we have, combining expressions (2.2.3), (2.3.4), and (2.3.5),

$$b_k(\theta) = 1 - e^{-\theta x} - \sum_{j=2}^{k} \frac{j^{j-2}(\theta x)^{j-1}e^{-j\theta x}}{(j-1)!}$$

$$= 1 - \sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{j-1}e^{-j\theta x}}{(j-1)!} \quad . \qquad (2.3.6)$$

This concludes the proof of Theorem 2.3. □

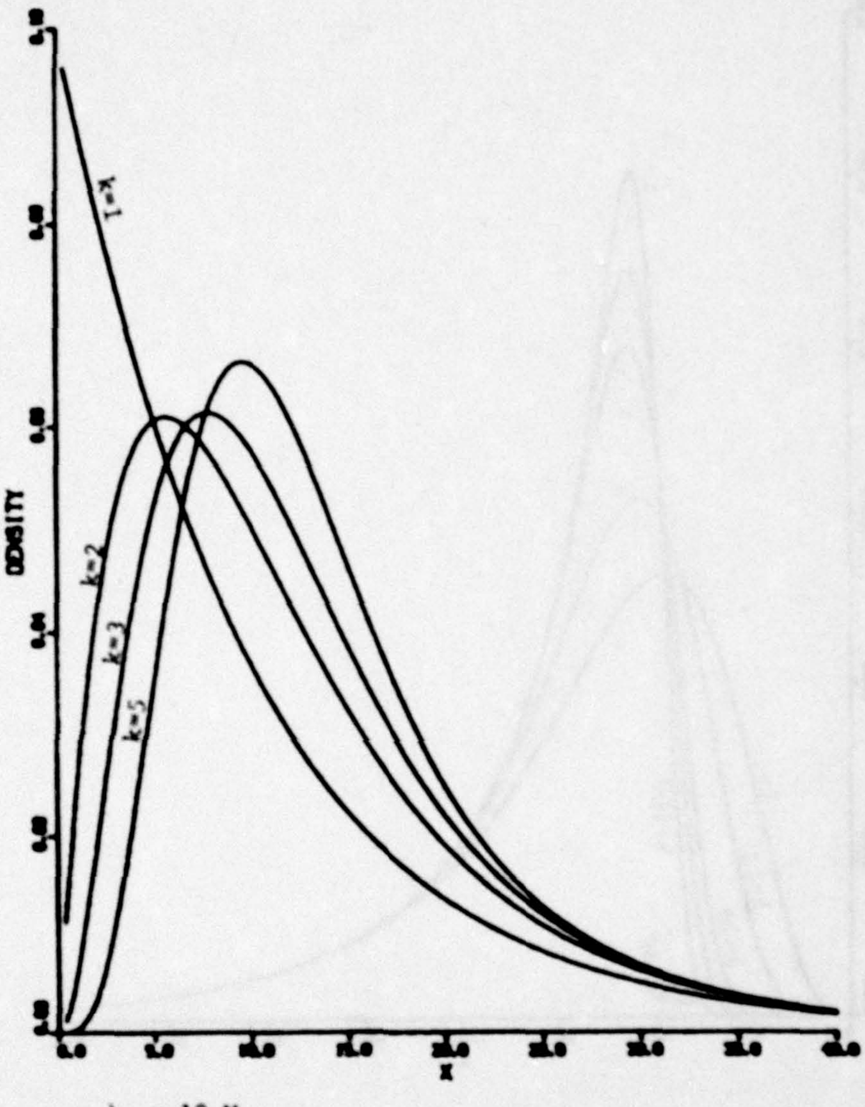
<u>Corollary</u>: The density function of $\hat{\lambda}_k$ in the equal λ case is given by

$$f_k(x) = \sum_{j=1}^{k} \frac{j^{j-2}}{(j-1)!} e^{-j\theta x} \theta(\theta x)^{j-2}[j\theta x - (j-1)] \ .$$

The corollary is obtained by differentiating expression (2.3.6) and combining terms.

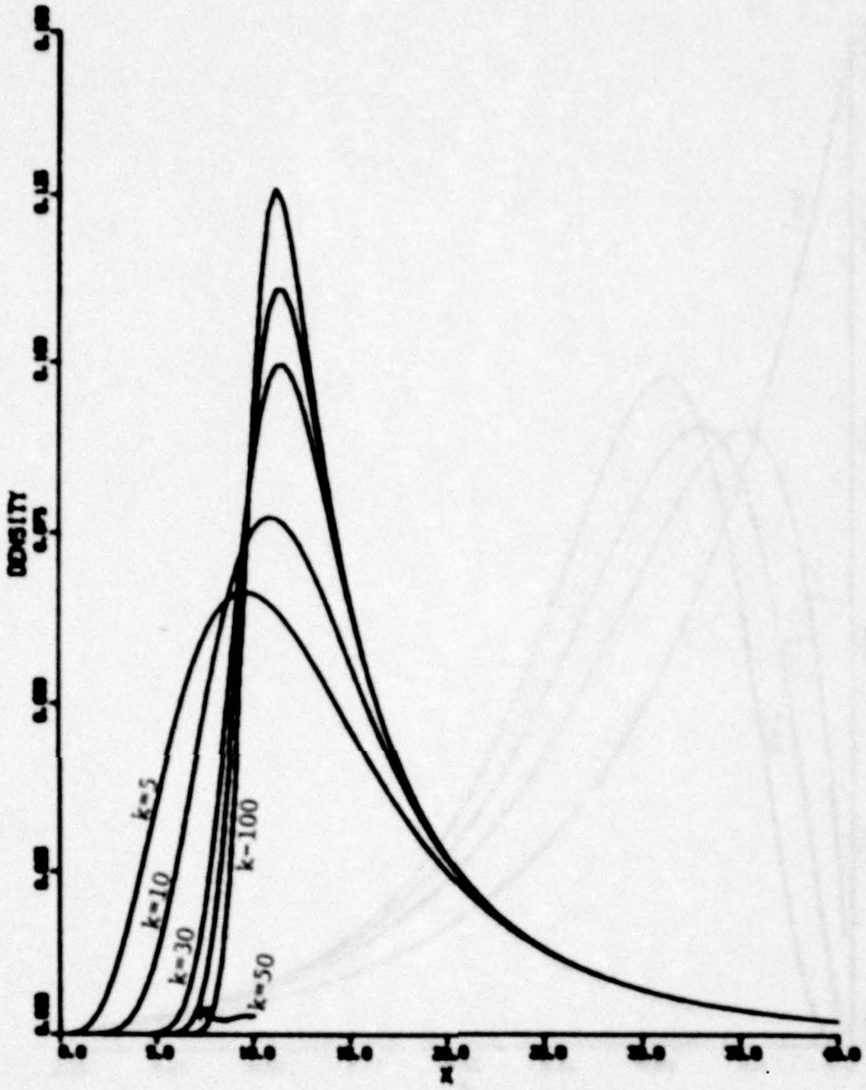
We now have expressions for the distribution and density functions of $\hat{\lambda}_k$ in the equal λ case for any value of k. Graphs of the density functions for $\lambda = 10$ and values of k ranging from 1 to 100 are given in Figures 1 and 2. Note that the densities have a heavy right-hand tail, which will be seen to lead to high variances for the $\hat{\lambda}_k$.

The moments of $\hat{\lambda}_k$ are easily found using expression (2.2.7) with $F_k(x)$ substituted for $F(x)$:



$\lambda_i = 10\ \forall_i$

Figure 1. Density of $\hat{\lambda}_k$ in the Case of Equal λ_i.



$\lambda_i = 10\ V_i$

Figure 2. Density of $\hat{\lambda}_i$ in the Case of Equal λ_i.

$$E(\hat{\lambda}_k^{\alpha}) = \alpha \int_0^{\infty} x^{\alpha-1} \left[\sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{j-1} e^{-j\theta x}}{(j-1)!} \right] dx$$

$$= \alpha \sum_{j=1}^{k} \frac{\theta^{j-1} j^{j-2}}{(j-1)!} \int_0^{\infty} x^{\alpha+j-2}\, e^{-j\theta x}\, dx$$

$$= \alpha \sum_{j=1}^{k} \frac{\theta^{j-1} j^{j-2}}{(j-1)!} \frac{1}{(j\theta)^{\alpha+j-1}} (\alpha + j - 2)!$$

$$= \frac{\alpha}{\theta^{\alpha}} \sum_{j=1}^{k} \frac{(\alpha+j-2)!}{(j-1)!\, j^{\alpha+1}}$$

$$= \alpha\, \lambda^{\alpha} \sum_{j=1}^{k} \frac{(\alpha+j-2)!}{(j-1)!\, j^{\alpha+1}} \,. \tag{2.3.7}$$

In particular, with $\alpha = 1$ and $\alpha = 2$ we obtain

$$E(\hat{\lambda}_k) = \lambda \sum_{j=1}^{k} \frac{1}{j^2} \equiv \lambda \eta(k) \tag{2.3.8}$$

$$E(\hat{\lambda}_k^{\,2}) = 2\lambda^2 \sum_{j=1}^{k} \frac{1}{j^2} = 2\lambda^2\, \eta(k) \tag{2.3.9}$$

so the first two moments are seen to depend on a p-series with $p = 2$. We denote the k^{th} partial sum of this series $\eta(k)$. Limiting properties (as $k \to \infty$) will be discussed in Section 2.6. Use expressions (2,3,8) and (2.3.9) to obtain

$$\mathrm{Var}(\hat{\lambda}_k) = 2\lambda^2\, \eta(k) - \lambda^2 \eta^2(k) \,.$$

Evaluating mean-square-error (MSE), which will be used in the future to evaluate the effectiveness of $\hat{\lambda}_k$ as an estimator, gives the interesting result

$$\begin{aligned} MSE(\hat{\lambda}_k) &= 2\lambda^2 \eta(k) - \lambda^2 \eta^2(k) + (\lambda\eta(k) - \lambda)^2 \\ &= 2\lambda^2 \eta(k) - \lambda^2 \eta^2(k) + \lambda^2 \eta^2(k) - 2\lambda^2 \eta(k) + \lambda^2 = \lambda^2 . \end{aligned}$$

So, in the case when the λ_i are all equal, the MSE of $\hat{\lambda}_k$ is independent of k, the number of stages.

2.4. Efficiency of Isotonic Estimators of the λ_i

The $\hat{\lambda}_i$ are intuitively attractive estimators of the mean lifetimes for two reasons. Firstly, they take the ordering of the λ_i into account by maximizing the likelihood function over the restricted space $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \cdots \leq \hat{\lambda}_k$. Secondly, each $\hat{\lambda}_i$ is a function of <u>all</u> of the random variables $X_1, X_2, \ldots, X_k$ and consequently each estimate makes use of all of the information available. This is desirable, for example, because although X_1 does not have mean λ_k, its mean is related to λ_k through the ordering relationship, and consequently X_1 may contain useful information regarding λ_k.

We wish to evaluate the performance of the isotonic estimates in the case of exponentially distributed X_i. The criteria for evaluation will be mean-square-error, a common measure of estimator effectiveness. As a basis for comparison, two competing estimates of

the λ_i are proposed. The first is to use $\bar{X}$, the overall mean of the X_i, as a common estimate of the λ_i. $\bar{X}$ would be an appropriate estimate in the boundary case where the λ_i are all equal (which makes the X_i iid random variables). In fact, $\bar{X}$ is easily seen to be the uniform minimum variance unbiased estimate (UMVUE) in this situation. At the beginning of the investigation, it was expected that $\bar{X}$ would perform comparatively well when the λ_i were equal or close to equal, but would not do as well when the λ_i varied widely.

The second proposed "competitor" of the $\hat{\lambda}_i$ is simply to use X_i as the estimate of λ_i at each stage. In effect, we treat the observations $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k$ as if they are totally unrelated to λ_i and consequently are of no value in estimating λ_i. It is interesting to note that if the λ_i are widely spaced (for example, if $\min \left\{\frac{\lambda_{i+1}}{\lambda_i}\right\}_{i=1}^{k-1}$ is approaching ∞), these estimates are effectively equal to the isotonic regression estimates, since the X_i are themselves non-decreasing with high probability. We would expect that the X_i estimates will do comparatively well when the λ_i are indeed widely spaced, but will not fare as well in other cases.

To compare the three proposed methods of estimation, two cases are examined. The first is total mean-square-error (MSE) over all the λ_i. In this case, the MSE for the $\bar{X}$ and X_i estimates are computed exactly while, due to computational difficulty, the MSE's for the $\hat{\lambda}_i$ are found via computer simulation. The second case is to restrict our investigation to the performance of estimators of the

final mean λ_k : $\hat{\lambda}_k$, $\bar{X}$, and X_k. Since we have developed a recursive technique for finding the distribution and moments of $\hat{\lambda}_k$ in the exponential case, all computations are exact and no simulation studies are required here.

The computation of MSE for the $\bar{X}$ and X_i estimates is straightforward. For $\bar{X}$ we have:

$$E(\bar{X}) = \frac{\sum_{j=1}^{k} \lambda_j}{k} \quad \text{and} \quad Var(\bar{X}) = \frac{\sum_{j=1}^{k} \lambda_j^2}{k^2} .$$

So, $MSE_i(\bar{X}) = Var(\bar{X}) + Bias_i^2(\bar{X})$

$$= \frac{\sum_{j=1}^{k} \lambda_j^2}{k^2} + \left(\frac{\sum_{j=1}^{k} \lambda_j}{k} - \lambda_i \right)^2 \qquad (2.4.1)$$

where $MSE_i(\bar{X})$ and $Bias_i(\bar{X})$ represent the mean-square-error and bias when $\bar{X}$ is used to estimate λ_i. For X_i (which is unbiased for λ_i) we have:

$$MSE(X_i) = Var(X_i) = \lambda_i^2 .$$

A computer program is used to compute the exact moments and MSE for $\hat{\lambda}_k$. Let us examine the fairly manageable case $k = 2$. $E(\hat{\lambda}_2)$ and $E(\hat{\lambda}_2^2)$ were given in Section 2.2, and $E(\hat{\lambda}_2)$ can be rewritten:

$$E(\hat{\lambda}_2) = \frac{1}{\theta_2} + \frac{\theta_2}{2\theta_1(\theta_1+\theta_2)} \quad .$$

Since θ_1 and θ_2 are both positive, this form of expression clearly shows the negative bias or tendency of $\hat{\lambda}_2$ to overestimate. This is true in general for all values of k, since

$$\hat{\lambda}_k = \max(X_k, \frac{X_k+X_{k-1}}{2}, \ldots, \bar{X}) \geq X_k$$

and X_k is unbiased for λ_k. Note that $\hat{\lambda}_k$ has a positive probability of being strictly greater than X_k in all but the trivial case $\lambda_1 = \lambda_2 = \cdots = \lambda_{k-1} = 0$. Now

$$\begin{aligned} MSE(\hat{\lambda}_2) &= Var(\hat{\lambda}_2) + Bias^2(\hat{\lambda}_2) \\ &= E(\hat{\lambda}_2^2) - E(\hat{\lambda}_2)^2 + (E(\hat{\lambda}_2) - \lambda_2)^2 \\ &= E(\hat{\lambda}_2^2) - 2\lambda_2 E(\hat{\lambda}_2) + \lambda_2^2 \ . \end{aligned}$$

Substituting the previously derived expressions for $E(\hat{\lambda}_2)$ and $E(\hat{\lambda}_2^2)$, we obtain, after some algebraic manipulation:

$$MSE(\hat{\lambda}_2) = \frac{1}{\theta_2^2} + \frac{(\theta_2-\theta_1)(2\theta_1+\theta_2)}{2\theta_1^2(\theta_1+\theta_2)^2} \ . \qquad (2.4.2)$$

Letting $k = 2$ and replacing λ_i by $1/\theta_i$ in expression (2.4.1), we obtain

$$MSE_2(\bar{X}) = \frac{\frac{1}{\theta_1^2} + \frac{1}{\theta_2^2}}{4} + \left(\frac{\frac{1}{\theta_1} + \frac{1}{\theta_2}}{2} - \frac{1}{\theta_2}\right)^2$$

which becomes, after algebraic manipulation:

$$MSE_2(\bar{X}) = \frac{1}{\theta_2^2} + \frac{1}{2\theta_1^2\theta_2^2}\left[-\theta_1^2 + \theta_2^2 - \theta_1\,\theta_2\right].$$

Let us compare $\hat{\lambda}_2$, $\bar{X}$, and X_2 as estimators of λ_2. This comparison will be done for values of (λ_1, λ_2) which satisfy the assumed relation $\lambda_1 \leq \lambda_2$ (or $\theta_1 \geq \theta_2$). Using this fact and the fact that $MSE(X_2) = Var(X_2) = \frac{1}{\theta_2^2}$, it is seen from expression (2.4.2) that:

$$MSE(X_2) \geq MSE(\hat{\lambda}_2) \qquad (2.4.3)$$

where equality holds only in the case where $\theta_1 = \theta_2$ (equal means). Now let us compare $MSE(\bar{X})$ and $MSE(\hat{\lambda}_2)$:

$$MSE(\hat{\lambda}_2) - MSE(\bar{X}) = \frac{(\theta_2-\theta_1)(2\theta_1+\theta_2)}{2\theta_1^2(\theta_1+\theta_2)^2} - \frac{1}{2\theta_1^2\theta_2^2}(-\theta_1^2 + \theta_2^2 - \theta_1\,\theta_2)$$

which after algebraic manipulation, is seen to be equal to:

$$\frac{\theta_1(\theta_1+3\theta_2)}{2\theta_2^2(\theta_1+\theta_2)^2}.$$

Under the restriction $\theta_1, \theta_2 > 0$ this term is always positive. Combining this with (2.4.3), we obtain the ordering

$$\mathrm{MSE}(X_2) \geq \mathrm{MSE}(\hat{\lambda}_2) > \mathrm{MSE}(\bar{X}) .$$

This holds for all pairs (λ_1, λ_2) where $\lambda_1 \leq \lambda_2$, and consequently we conclude that both X_2 and $\hat{\lambda}_2$ are inadmissible (with respect to MSE) in the region of interest. To illustrate this inadmissibility, Figure 3 gives a graph of $\mathrm{MSE}(X_2)$, $\mathrm{MSE}(\hat{\lambda}_2)$, and $\mathrm{MSE}(\bar{X})$ for a fixed value of λ_2 and values of λ_1 ranging from 0 to λ_2. Note that $\hat{\lambda}_2$ is only a slight improvement over X_2, while $\bar{X}$ is far superior to either X_2 or $\hat{\lambda}_2$.

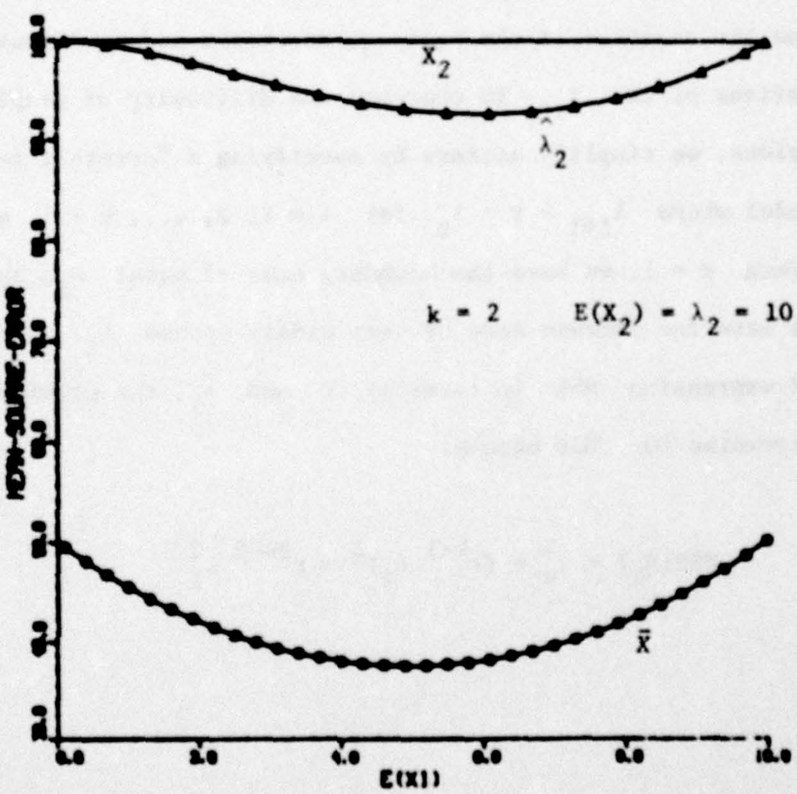


Figure 3. Mean-Square-Error Comparison for the Case $k = 2$.

The case where $k = 2$ and we consider the estimation of λ_2 alone is the only case where an ordering relation among the mean-square-errors has been proven (note we have seen that $MSE(X_k) = MSE(\hat{\lambda}_k) = \lambda_k^2$ for any value of k where $\lambda_1 = \lambda_2 = \cdots = \lambda_k$). However, upon calculation of mean-square-errors for many values of k and for various configurations of the λ_i (assuming $\lambda_1 \leq \cdots \leq \lambda_k$), no case has been observed where the relation

$$MSE(X_k) \geq MSE(\hat{\lambda}_k) > MSE(\bar{X})$$

did not hold. Consequently, it is hypothesized that this relationship is true for all values of k.

As an illustration we consider the case $k = 5$, and would like to examine the relation of the mean-square-errors under various configurations of the λ_i. To overcome the difficulty of graphing in 6 dimensions, we simplify matters by specifying a "constant improvement ratio" model where $\lambda_{i+1} = r \cdot \lambda_i$ for $i = 1, 2, \ldots, k-1$ and $r \geq 1$. When $r = 1$, we have the boundary case of equal λ_i, and as $r \to \infty$ we have the extreme case of very widely spaced λ_i. Using this model and expressing MSE in terms of r and λ_i, the previously derived formulas for MSE become:

$$MSE(X_k) = \lambda_k^2 = (r^{k-1}\lambda_1)^2 = r^{2k-2}\lambda_1^2$$

and

$$\mathrm{MSE}_k(\bar{X}) = \frac{\sum_{i=1}^{k} r^{2i-2}\lambda_1^2}{k^2} + \left(\frac{\sum_{i=1}^{k} r^{i-1}\lambda_1}{k} - \lambda_k\right)^2$$

$$= \frac{\lambda_1^2}{k^2}\left(\frac{r^{2k}-1}{r^2-1}\right) + \left(\frac{\lambda_1}{k}\left(\frac{r^k-1}{r-1}\right) - r^{k-1}\lambda_1\right)^2 .$$

As before, no general closed formula for $\mathrm{MSE}(\hat{\lambda}_k)$ has been found. Figure 4 gives a MSE comparison of the three estimates of λ_k for values of r ranging from 1 to ∞. The limiting value for $\mathrm{MSE}(\hat{\lambda}_5)/\mathrm{MSE}_5(\bar{X})$ is found by noting that $\lim_{r\to\infty} P(\hat{\lambda}_5 = X_5) = 1$ and evaluating $\lim_{r\to\infty} \mathrm{MSE}(X_5)/\mathrm{MSE}(\bar{X})$. Dividing the MSE formulas given above, we obtain (for general k):

$$\lim_{r\to\infty} \frac{\mathrm{MSE}(X_k)}{\mathrm{MSE}(\bar{X})} = \frac{1}{\frac{1}{k^2} + (\frac{1}{k} - 1)^2} = \frac{k^2}{k^2 - 2k+2} .$$

Note from Figure 4 that $\bar{X}$ has considerably better MSE than either X_k or $\hat{\lambda}_k$ for all values of r, and that $\hat{\lambda}_k$ gives little improvement over X_k, particularly when r is large. This is not surprising since examination of expression (1.3.3) shows that $\hat{\lambda}_k$ will actually be equal to X_k with increasing frequency as $r \to \infty$.

In an attempt to determine why the isotonic estimator performs so poorly, the MSE is divided into its components of variance and bias. Table 1 gives the variance and expectation of $\hat{\lambda}_k$

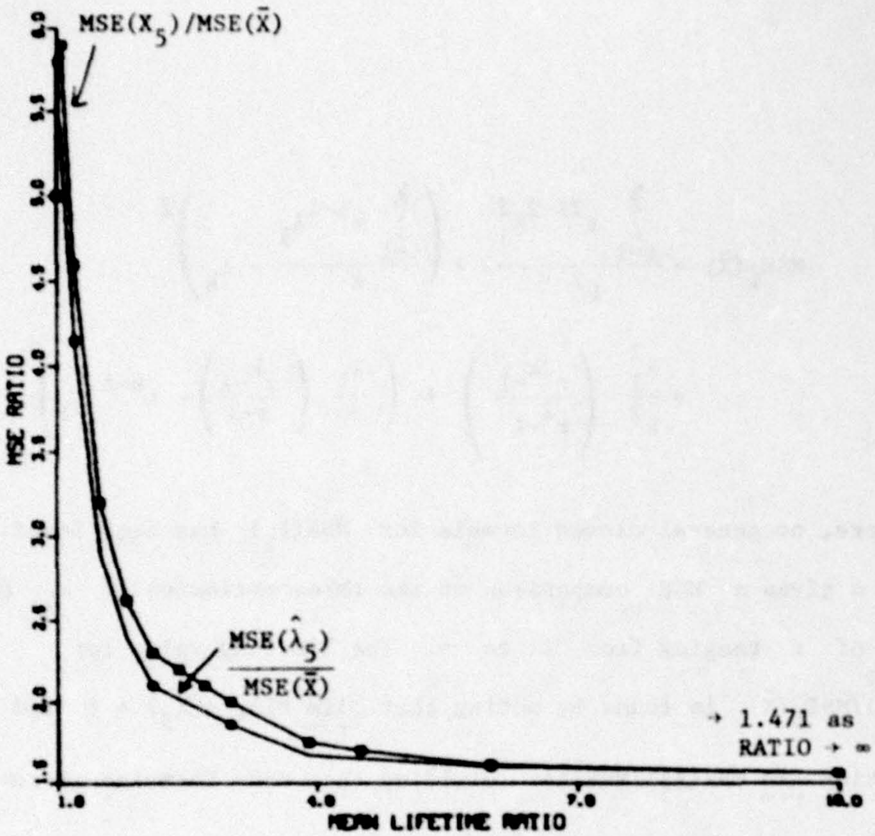
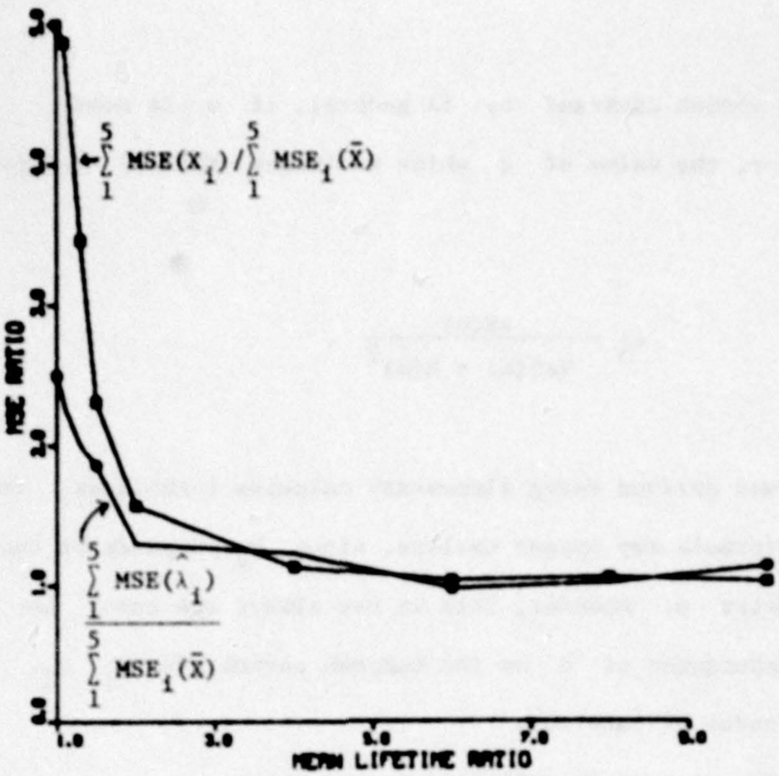


Figure 4. Mean-Square-Error Comparison for the Case $k = 5$.

(in the case $k = 5$) for various values of r. As the table shows, $\hat{\lambda}_5$ is competitive with $\bar{X}$ with respect to bias (becoming significantly better as r increases), but not at all competitive with respect to variance. Empirically, $\hat{\lambda}_5$ gives a maximum reduction in variance over the use of X_5 alone of about 20%, while $\bar{X}$ can be shown to reduce this variance as much as 96% $(1-1/k^2)$ when $r \to \infty$. Consequently, even though $\bar{X}$ is heavily biased for large values of r, its low variance more than compensates to give a reasonable MSE.

So far, we have only compared mean-square-errors for the estimates of λ_k alone. Now consider total MSE over all k of the λ_i. Figure 5 gives a comparison similar to that of Figure 4,



$k = 5$

Figure 5. Comparison of Total MSE for Three Proposed Estimators

replacing final estimate MSE's by totals. For most reasonable improvement ratios, $\bar{X}$ is seen to have the best total MSE. For large improvement ratios, starting in the neighborhood of 6, the three estimation techniques are roughly equivalent.

2.5. A Simple Modification of the Estimators of λ_k

The initial performance of the isotonic estimator $\hat{\lambda}_k$ in the exponential case is somewhat disappointing. We wish to determine if the situation can be improved through simple modification of the proposed estimators. Consider multiplying an estimator by an

appropriately chosen constant c. In general, if $\hat{\alpha}$ is some estimator of α, the value of c which minimizes $MSE(c\hat{\alpha})$ is given by

$$c_{\hat{\alpha}} = \frac{\alpha E(\hat{\alpha})}{Var(\hat{\alpha}) + E(\hat{\alpha})^2} .$$

This formula was derived using elementary calculus techniques. In practice the formula may appear useless, since $c_{\hat{\alpha}}$ depends on the unknown parameter α. However, this is not always the case. Let us examine the dependence of c on the unknown parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ in the three cases of interest.

When X_k is used to estimate λ_k, we obtain

$$c_{X_k} = \frac{\lambda_k(\lambda_k)}{\lambda_k^2 + \lambda_k^2} = \frac{1}{2} .$$

So c_{X_k} is independent of λ_k, and we can always assure a reduction in MSE by dividing X_k by 2.

When $\bar{X}$ is used to estimate λ_k, we obtain

$$c_{\bar{X}} = \frac{\lambda_k\left(\frac{\sum_1^k \lambda_i}{k}\right)}{\frac{\sum_1^k \lambda_i^2}{k^2} + \left(\frac{\sum_1^k \lambda_i}{k}\right)^2} = \frac{k\lambda_k \sum_1^k \lambda_i}{\sum_1^k \lambda_i^2 + \left(\sum_1^k \lambda_i\right)^2} .$$

Note that $c_{\bar{X}}$ does depend on the unknown λ_i, and consequently cannot be determined. In the case of equal λ_i, this does reduce to $\frac{k}{k+1}$, which is independent of the λ_i.

Now consider $\hat{\lambda}_k$. As with $\bar{X}$, $c_{\hat{\lambda}_k}$ will be a function of the λ_i. We will compute $c_{\hat{\lambda}_k}$ for the two "extreme" cases of equal λ_i and widely spaced λ_i. In the equal λ_i case, we substitute the moment formulas derived in Section 2.3 to obtain

$$c_{\hat{\lambda}_k} = \lambda(\eta(k))/(2\lambda^2\ \eta(k)) = 1/2 .$$

In the case of widely spaced λ_i (for example, consider the constant ratio model with $r \to \infty$), $\hat{\lambda}_k$ is effectively equal to X_k, and we obtain:

$$c_{\hat{\lambda}_k} = c_{x_k} = 1/2 .$$

So we know that $c_{\hat{\lambda}_k}$ is equal to the same value at both "extreme" configurations of the λ_i. This leads to questioning the sensitivity of $c_{\hat{\lambda}_k}$ to changes in the λ_i. Empirical investigation shows that $c_{\hat{\lambda}_k}$ is very insensitive to the values of the λ_i. To illustrate this, consider the familiar "constant improvement ratio" model with $k = 5$. Figure 6 shows a graph of $c_{\hat{\lambda}_k}$ vs. r. The consistency of the results leads one to feel confident that a nearly optimal improvement would occur if $\hat{\lambda}_k$ were divided by 2.

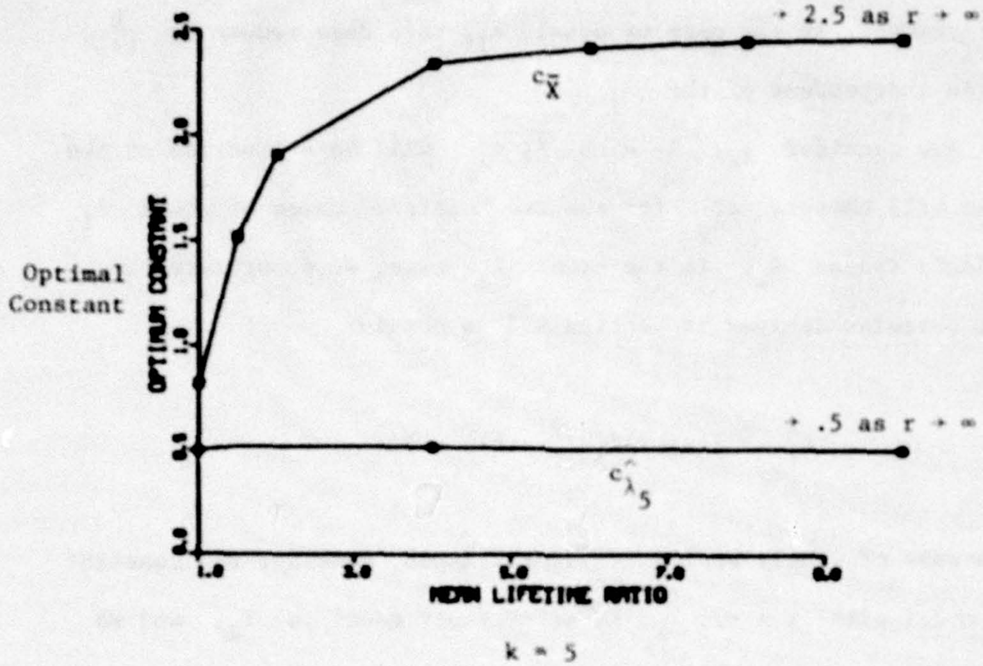
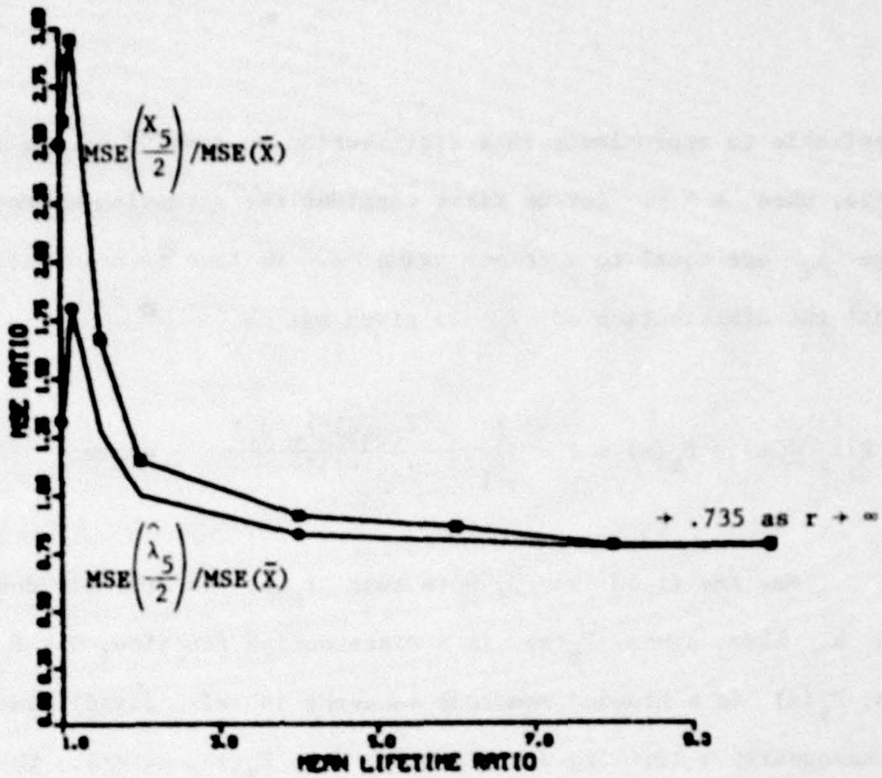


Figure 6. Optimal Multiplicity Constants

Note that Figure 6 also shows the relationship of $c_{\bar{X}}$ to r. In this case, $c_{\bar{X}}$ is definitely sensitive to r (or to the λ_i). Consequently, we cannot confidently state that $\bar{X}$ will be improved by using a particular multiplicative constant unless we know more about the λ_i.

Figure 7 illustrates the relationship among mean-square-errors for the estimators $X_k/2$, $\hat{\lambda}_k/2$, and $\bar{X}$. $\bar{X}$ is left unmodified due to the previously discussed lack of information regarding the proper $c_{\bar{X}}$. As the graph shows, we have finally found situations where an isotonic-based estimate is worth using in the sense of having reasonable



$k = 5$

Figure 7. Relative Mean-Square-Errors for Modified Estimators

mean-square-error. However, we do require a large improvement in mean lifetime at each stage (r must be somewhat larger than 2 in this example) in order to realize the benefit of using the isotonic estimate instead of the overall mean $\bar{X}$. Also, there is no substantial improvement when using $\hat{\lambda}_5/2$ rather than $X_5/2$ in the cases where $\hat{\lambda}_5/2$ did outperform $\bar{X}$.

2.6. Asymptotic Distribution of $\hat{\lambda}_k$ as $k \rightarrow \infty$ (Exponential Case)

For large values of k, computation of the distribution of $\hat{\lambda}_k$ may be time-consuming and awkward. Consequently, it would be

desirable to approximate this distribution by looking at the limiting case, when $k \to \infty$. Let us first consider the situation where all of the λ_i are equal to a common value λ. We have seen in Section 2.3 that the distribution of $\hat{\lambda}_k$ is given by:

$$P\{\hat{\lambda}_k \leq x\} = F_k(x) = 1 - \sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{j-1} e^{-j\theta x}}{(j-1)!} \qquad \text{where} \qquad \theta = 1/\lambda .$$

For any fixed $x > 0$, note that $F_k(x)$ is strictly decreasing in k. Also, since $F_k(x)$ is a distribution function, $0 \leq F_k(x) \leq 1$. So, $F_k(x)$ is a bounded monotone sequence in k(x fixed) and consequently a limiting distribution, say, $F_\infty(x)$, exists. Therefore, we can say that $\hat{\lambda}_k$ converges weakly or in distribution to a limiting random variable $\hat{\lambda}_\infty$ having distribution function $F_\infty(x)$. Moments of $\hat{\lambda}_\infty$ are found as follows:

$$E(\hat{\lambda}_\infty^\alpha) = \alpha \int_0^\infty x^{\alpha-1} [1 - F_\infty(x)]\, dx$$

$$= \alpha \int_0^\infty x^{\alpha-1} \left[\lim_{k\to\infty} \sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{j-1} e^{-j\theta x}}{(j-1)!} \right] dx$$

Since the summation inside the integral is monotone in k, we can use a monotone convergence theorem (cf. Feller, Vol. II, page 110 [9]) to interchange limit and integral signs and obtain

$$E(\hat{\lambda}_\infty^\alpha) = \lim_{k\to\infty} \alpha \int_0^\infty x^{\alpha-1} \sum_{j=1}^{k} \frac{j^{j-2}(\theta x)^{j-1} e^{-j\theta x}}{(j-1)!} .$$

The integral given above was evaluated in expression (2.3.7), and substituting that result we obtain

$$E(\hat{\lambda}_\infty^\alpha) = \lim_{k\to\infty} \alpha\lambda^\alpha \sum_{j=1}^{k} \frac{\Gamma(\alpha+j-1)}{\Gamma(j)j^{\alpha+1}}$$

$$= \alpha\lambda^\alpha \sum_{j=1}^{\infty} \frac{\Gamma(\alpha+j-1)}{\Gamma(j)j^{\alpha+1}} .$$

In particular, with $\alpha = 1$ and $\alpha = 2$ we have

$$E(\hat{\lambda}_\infty) = \lambda \sum_{j=1}^{\infty} 1/j^2 = \lambda\eta(\infty) \approx 1.644\ \lambda$$

$$E(\hat{\lambda}_\infty^2) = 2\lambda^2 \sum_{j=1}^{\infty} 1/j^2 = 2\lambda^2\ \eta(\infty) \approx 3.288\ \lambda^2 .$$

So, $\mathrm{Var}(\hat{\lambda}_\infty) \approx .585\ \lambda^2$ and $\mathrm{MSE}(\hat{\lambda}_\infty) = \lambda^2$. These results illustrate another difficulty in the use of the isotonic estimation procedure. By the Kolmogorov strong law of large numbers, $\bar{X}$ converges almost surely to λ as $k \to \infty$, and $\mathrm{MSE}(\bar{X}) \to 0$. However, $\hat{\lambda}_k$ has the unfortunate characteristic of maintaining a constant non-zero MSE while k increases and more information about λ becomes available.

The asymptotic properties derived above may seem to be of little value in that the assumption of equal λ_i is likely to be unrealistic. In general, of course, the asymptotic distribution of $\hat{\lambda}_k$ will depend on the configuration of the λ_i. With our prototype development model, it would be reasonable to assume that we are approaching some optimal level of design as stagewise improvements

are made. Consequently, let us consider the case where the λ_i converge upward to a limiting value λ_∞. The following theorem allows us to equate this situation with the previously considered case of equal means.

Theorem 2.4: The asymptotic distribution of $\hat{\lambda}_k$, the isotonic estimate of the final mean λ_k, is identical in the following two cases:

(1) the X_i are independent exponentially distributed random variables having common mean λ;

(2) the X_i are independent exponentially distributed random variables having individual means, where $\lambda_i \uparrow \lambda$.

Proof: Assume $X_i \sim \exp(\lambda_i)$ where $\lambda_i \uparrow \lambda$. Let $\theta_i = 1/\lambda_i$ and $\theta = 1/\lambda$. Let $Y_i = (\theta_i/\theta)X_i$, so that the Y_i are independent exponentially distributed random variables with common mean λ. Now define

$$A_k = \max(X_k, \frac{X_k+X_{k-1}}{2}, \ldots, \bar{X})$$

$$B_k = \max(Y_k, \frac{Y_k+Y_{k-1}}{2}, \ldots, \bar{Y})$$

so A_k and B_k are the isotonic estimates based on the X_i and Y_i, respectively. Since the Y_i have common mean λ, we know B_k converges in distribution to a limiting random variable $\hat{\lambda}_\infty$ with distribution function

$$1 - \sum_{j=1}^{\infty} \frac{j^{j-2}(\theta x)^{j-1} e^{-j\theta x}}{(j-1)!} .$$

We wish to show that $B_k - A_k \xrightarrow{P} 0$. (Note that $B_k \geq A_k$ since $\theta_i/\theta \geq 1 \; \forall_i$.). To do this, we need to show for any value δ that $P\{B_k - A_k > \delta\} \to 0$ as $k \to \infty$. Equivalently, we must show for any δ:

$$\forall \varepsilon > 0 \; \exists N \ni k \geq N \Rightarrow P\{B_k - A_k > \delta\} < \varepsilon .$$

Use the following inequality:

$$\max(b_1, b_2, \ldots, b_k) - \max(a_1, a_2, \ldots, a_k) \leq \max(b_1 - a_1, \ldots, b_k - a_k)$$

to obtain

$$B_k - A_k \leq \max\left(Y_k - X_k, \frac{\left(Y_k + Y_{k-1}\right) - \left(X_k + X_{k-1}\right)}{2}, \ldots, \bar{Y} - \bar{X}\right)$$

$$= \max\left(X_k\left(\frac{\theta_k}{\theta} - 1\right), \frac{X_k\left(\frac{\theta_k}{\theta} - 1\right) + X_{k-1}\left(\frac{\theta_{k-1}}{\theta} - 1\right)}{2}, \ldots, \frac{X_k\left(\frac{\theta_k}{\theta} - 1\right) + \cdots + X_1\left(\frac{\theta_1}{\theta} - 1\right)}{k}\right).$$

Now define $\alpha = \frac{\varepsilon\delta}{8\lambda}$. Since $\frac{\theta_i}{\theta} \downarrow 1$, we know that:

$$\exists\, n \ni i \geq n \Rightarrow \frac{\theta_i}{\theta} - 1 < \alpha .$$

Therefore, for $i \geq n$ we have, replacing $\frac{\theta_i}{\theta} - 1$ by α:

$$B_k - A_k \leq \max\left(\alpha X_k, \alpha\left(\frac{X_k + X_{k-1}}{2}\right), \ldots, \alpha\left(\frac{X_k + X_{k-1} + \cdots + X_n}{k-n+1}\right)\right),$$

$$\frac{\alpha\left(X_k + \cdots + X_n\right) + \left(\frac{\theta_{n-1}}{\theta} - 1\right) X_{n-1}}{k-n+2}, \ \ldots,$$

$$\left.\frac{\alpha\left(X_k + \cdots + X_n\right) + \left(\frac{\theta_{n-1}}{\theta} - 1\right)X_{n-1} + \cdots + \left(\frac{\theta_1}{\theta} - 1\right)X_1}{k}\right).$$

Now consider the inequality:

$$\max(a_1 + b_1, \ \ldots, \ a_k + b_k) \leq \max(a_1, \ \ldots, \ a_k) + \max(b_1, \ \ldots, \ b_k)$$

with the following values substituted for the a_i and b_i:

$$a_1 = \alpha X_k \qquad\qquad b_1 = b_2 = \cdots = b_{k-n+1} = 0$$

$$a_2 = \alpha\left(\frac{X_2 + X_{k-1}}{2}\right) \qquad\qquad b_{k-n+2} = \frac{\left(\frac{\theta_{n-1}}{\theta} - 1\right)X_{n-1}}{k-n+2}$$

$$\vdots \qquad\qquad \vdots$$

$$a_{k-n+1} = \alpha\left(\frac{X_k + \cdots + X_n}{k-n+1}\right) \qquad\qquad b_k = \frac{\left(\frac{\theta_{n-1}}{\theta} - 1\right) X_{n-1} + \cdots + \left(\frac{\theta_1}{\theta} - 1\right)X_1}{k}$$

$$a_{k-n+2} = \alpha\left(\frac{X_k + \cdots + X_n}{k-n+2}\right)$$

$$\vdots$$

$$a_k = \alpha\left(\frac{X_k + \cdots + X_n}{k}\right)$$

Using this inequality and the fact that $\frac{\theta_1}{\theta} - 1 \geq \frac{\theta_i}{\theta} - 1 \quad \forall i$ (since the θ_i are decreasing), we obtain:

$$B_k - A_k \le \max\left[\alpha X_k, \frac{\alpha(X_k+X_{k-1})}{2}, \ldots, \frac{\alpha(X_k+\cdots+X_n)}{k-n+1}, \ldots, \frac{\alpha(X_k+\cdots+X_n)}{k}\right]$$
$$+ \max\left[\frac{\left(\frac{\theta_1}{\theta}-1\right)X_{n-1}}{k-n+2}, \ldots, \frac{\left(\frac{\theta_1}{\theta}-1\right)(X_{n-1}+\cdots+X_1)}{k}\right].$$

Since $0 \le X_i \le Y_i \;\forall i$, it is easily seen (via a term-by-term comparison) that the first term on the right-hand-side is less than or equal to αB_k, so we obtain:

$$B_k - A_k \le \alpha B_k + \left(\frac{\theta_1}{\theta}-1\right)\max\left(\frac{X_{n-1}}{k-n+2}, \ldots, \frac{X_{n-1}+\cdots+X_1}{k}\right) \equiv \alpha B_k + Z_k .$$

Since n is fixed, it is easily seen that $Z_k \xrightarrow{a.s.} 0$ as $k \to \infty$, and consequently:

$$\exists m \ni k \ge m \Rightarrow P\{Z_k > \delta/\alpha\} < \varepsilon/2 .$$

Now, using an inequality similar to Chebychev's, we have the result:

$$P\{\alpha B_k > \delta/2\} \le \frac{E(\alpha B_k)}{\delta/2} = \frac{2\alpha E(B_k)}{\delta} .$$

Now, $E(B_k) = \lambda \sum_1^k 1/j^2 \le \lambda \sum_1^\infty 1/j^2 \le 2\lambda$, so we have:

$$P\{\alpha B_k > \delta/2\} \le \frac{2\alpha(2\lambda)}{\delta} .$$

Recall that $\alpha = \frac{\varepsilon\delta}{8\lambda}$, so we obtain:

$$P\{\alpha B_k > \delta/2\} \le \frac{4\lambda}{\delta}\left(\frac{\varepsilon\delta}{8\delta}\right)\frac{\varepsilon}{2} = \frac{\varepsilon}{2} .$$

Let $N = \max(n,m)$. By manipulating probabilities and using results given above we obtain, for $k \geq N$:

$$P\{B_k - A_k > \delta\} \le P\{\alpha B_k + Z_k > \delta\} \le P\{[\alpha B_k > \delta/2] \cup [Z_k > \delta/2]\}$$

$$\le P\{\alpha B_k > \delta/2\} + P\{Z_k > \delta/2\} \le \varepsilon/2 + \varepsilon/2 = \varepsilon .$$

Thus we have shown, for given values of δ and ε:

$$\exists N = \max(n,m) \ni k \geq N \Rightarrow P\{B_k - A_k > \delta\} < \varepsilon .$$

Therefore, by definition, $B_k - A_k \xrightarrow{P} 0$. Now recall Stutsky's theorem, which states that if $W_k \xrightarrow{d} W_\infty$ and if $W_k - Z_k \xrightarrow{P} 0$, then $Z_k \xrightarrow{d} W_\infty$. Using this theorem, we obtain the result that $A_k \xrightarrow{d} \hat{\lambda}_\infty$ so A_k and B_k have the same asymptotic distribution. □

The theorem serves to show that the asymptotic distribution derived for the equal λ_i case is also applicable in other situations. Figure 8 shows the distribution function $F_k(x)$ for various large values of k. Note that convergence to the asymptotic distribution, $F_\infty(x)$, is extremely slow. Consequently, use of an asymptotic approximation for finite but "large" values of k is to be approached with caution.

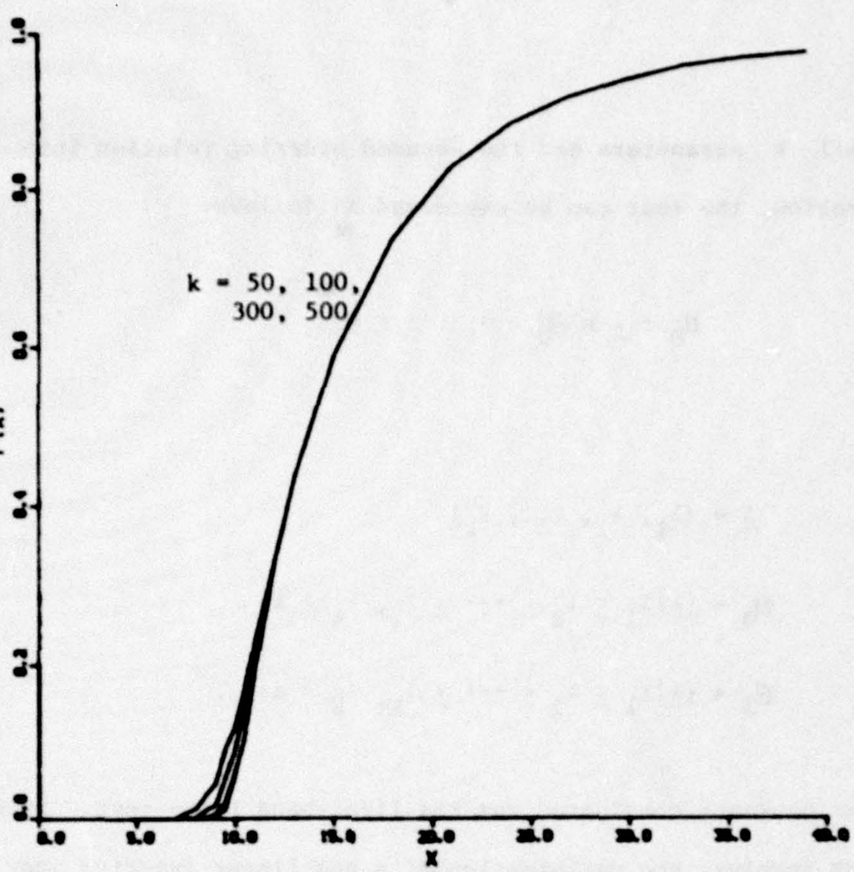


Figure 8. $F_k(x)$ for Large Values of k

2.7. Hypothesis Testing in the Exponential Case

In previous sections, we have discussed and evaluated various techniques for obtaining point estimates of λ_k, the final or most recent mean lifetime in the prototype development process. Let us now consider the related problem of testing hypotheses about the value of λ_k. First consider testing the hypothesis that a given mean lifetime a has been attained, versus the alternative that it has not:

$$H_0 : \lambda_k \geq a \quad \text{vs.} \quad H_1 : \lambda_k < a .$$

Taking all k parameters and the assumed ordering relation into consideration, the test can be expressed as follows:

$$H_0 : \underline{\lambda} \in \Theta_0 \quad H_1 : \underline{\lambda} \in \Theta_1$$

where:

$$\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$$

$$\Theta_0 = \{\underline{\lambda} | \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k, \lambda_k \geq a\}$$

$$\Theta_1 = \{\underline{\lambda} | \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k, \lambda_k < a\} \quad .$$

The first approach considered was the likelihood ratio test. Since this test involves the maximization of a non-linear function over a restricted space, it is somewhat awkward computationally and consequently we wish to examine some simpler alternatives. Consider tests based on the three previously developed point estimates of λ_k : $\hat{\lambda}_k$, $\bar{X}$, and X_k. Intuitively reasonable tests would be those which reject H_0 if (I) $\hat{\lambda}_k < c_1$; (II) $\bar{X} < c_2$; or (III) $X_k < c_3$, for appropriately chosen constants c_1, c_2, and c_3.

To obtain a size α test based on $\hat{\lambda}_k$, c_1 must be chosen so that:

$$\sup_{\Theta_0} P\{\hat{\lambda}_k < c_1\} = \alpha \quad .$$

It is easily seen that this rejection probability is maximized over Θ_0 when $\lambda_1 = \lambda_2 = \cdots = \lambda_{k-1} = 0$ and $\lambda_k = a$. With this choice of $\underline{\lambda}$ we have $\hat{\lambda}_k = X_k$ and consequently c_1 is chosen so that

$$P_{\lambda_k = a}\{X_k < c_1\} = \alpha$$

or

$$1 - e^{-c_1/a} = \alpha$$

or

$$c_1 = -a(\ln(1-\alpha)) \ .$$

So, the test based on $\hat{\lambda}_k$ rejects if $\hat{\lambda}_k < -a(\ln(1-\alpha))$.

Now consider a size α test based on $\bar{X}$ or, equivalently, $\sum_1^k X_i$. In this case we choose a value c_4 so that

$$\sup_{\Theta_0} P\left\{\sum_1^k X_i < c_4\right\} = \alpha \ .$$

As in the previous test, this rejection probability is seen to be maximized when $\lambda_1 = \cdots = \lambda_{k-1} = 0$ and $\lambda_k = a$. For this value of $\underline{\lambda}$, $\sum_1^k X_i = X_k$, so the evaluation of c_4 is identical to that of c_1 in the previous test and the test rejects H_0 if $\sum_1^k X_i < -a(\ln(1-\alpha))$.

It is now easily seen that the third proposed test, based on X_k, would reject if $X_k < -a(\ln(1-\alpha))$.

The three critical values found above all are equal because the three test statistics $(\hat{\lambda}_k, \sum_1^k X_i$, and $X_k)$ are equal for the

common value of $\underline{\lambda}$ at which the supremum rejection probabilities are attained. Now it is easily seen that the ordering:

$$X_k \leq \hat{\lambda}_k \leq \sum_1^k X_i$$

holds since the X_i are non-negative. Letting $c = -a(\ln(1-\alpha))$, the common critical value, we obtain the following ordering among the power functions:

$$P_{\underline{\lambda}}\left\{X_k < c\right\} \geq P_{\underline{\lambda}}\left\{\hat{\lambda}_k < c\right\} \geq P_{\underline{\lambda}}\left\{\sum_1^k X_i < c\right\}$$

for any value $\underline{\lambda} \in \Theta_0 \cup \Theta_1$. In this case there is no "best" test. If $\underline{\lambda} \in \Theta_0$, the test based on $\sum_1^k X_i$ is most desirable since it is least likely to falsely reject the null hypothesis. However, if $\underline{\lambda} \in \Theta_1$, the test based on X_k is most desirable since it is most likely to correctly reject the null hypothesis. The choice of a test in this situation would depend on what kind of error the user is most anxious to guard against - that of falsely rejecting a valid null hypothesis or falsely accepting an invalid one. Figure 9 gives examples of the behavior of the three tests when testing $H_0 : \lambda_3 \geq 10$ vs. $H_1 : \lambda_3 < 10$ at the $\alpha = .05$ level when $k = 3$. Note that the tests involving $\hat{\lambda}_3$ and $\sum_1^k X_i$ are very biased - that is, there are values of $\underline{\lambda}$ in the Θ_1 region for which the probability of accepting H_0 is greater than $1 - \alpha$. This undesirable

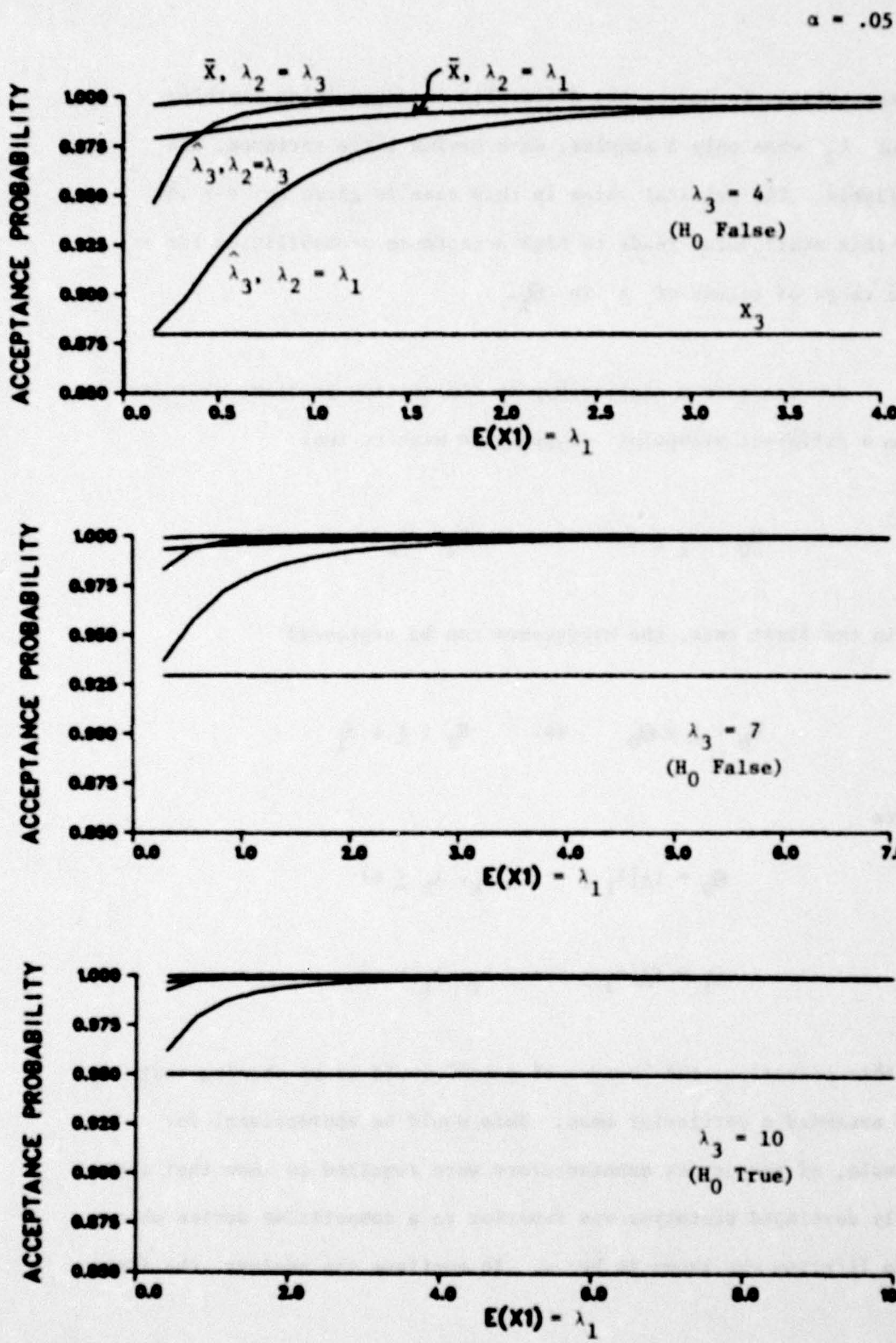


Figure 9. Comparative Behavior of Tests for $H_0 : \lambda_3 \geq 10$

characteristic is due to the difficulty in determining anything about λ_3 when only 3 samples, each having large variance, are available. The critical value in this case is given by $c = .51$, and this small value leads to high acceptance probabilities for a wide range of values of $\underline{\lambda}$ in Θ_1.

Now consider a similar hypothesis testing problem, expressed from a different viewpoint. Suppose we wish to test:

$$H_0 : \lambda_k \leq a \qquad \text{vs.} \qquad H_1 : \lambda_k > a \ .$$

As in the first test, the hypotheses can be expressed:

$$H_0 : \underline{\lambda} \in \Theta_0 \qquad \text{vs.} \qquad H_1 : \underline{\lambda} \in \Theta_1$$

where

$$\Theta_0 = \{\underline{\lambda} | \lambda_1 \leq \cdots \leq \lambda_k, \ \lambda_k \leq a\}$$

$$\Theta_1 = \{\underline{\lambda} | \lambda_1 \leq \cdots \leq \lambda_k, \ \lambda_k > a\} \ .$$

In this situation, the "burden of proof" would be on showing that λ_k had exceeded a particular mean. This would be appropriate, for example, if the device manufacturers were required to show that their newly developed prototype was superior to a competitive device whose mean lifetime was known to be a. To continue the analogy, the first

testing problem ($H_0 : \lambda_k \geq a$ vs. $H_1 : \lambda_k < a$) presented would be appropriate if the competitor were required to prove that the newly developed prototype was inferior to his own device.

Using the same logic as in the original testing situation, intuitively reasonable tests would be those which reject H_0 if (I) $\hat{\lambda}_k > c_1$; (II) $\sum_1^k X_i > c_2$; or (III) $X_k > c_3$. As before, c_1, c_2, and c_3 are constants chosen to give the tests the required size.

In the first case, we obtain a size-α test if c_1 is chosen so that:

$$\sup_{\Theta_0} P\{\hat{\lambda}_k > c_1\} = \alpha .$$

This rejection probability is maximized over Θ_0 when $\lambda_1 = \lambda_2 = \cdots = \lambda_k = a$. Letting $\theta_a = 1/a$ and using the equal-mean distribution developed in Section 2.3, we obtain the equation:

$$1 - F_k(c_1) = \alpha \qquad \text{or} \qquad F_k(c_1) = 1 - \alpha$$

where

$$F_k(c_1) = 1 - \sum_{j=1}^{k} \frac{j^{j-2}(\theta_a c_1)^{j-1} e^{-j\theta_a c_1}}{(j-1)!} .$$

So, it is seen that the critical value, c_1, is equal to the appropriate percentage point of the equal-mean distribution. This distribution

depends only on the product $\theta_a c_1$, which greatly simplifies its tabulation. Table 2 gives the $(1 - \alpha)$th percentage points for various values of α and k. Note that convergence has been reached by $k = 20$, so the table is appropriate for all values of k.

Suppose, for example, that $k = 5$ and we wish to test

$$H_0 : \lambda_5 \leq 10 \quad \text{vs.} \quad H_1 : \lambda_5 > 10$$

using $\alpha = .05$. From the table, the appropriate value for θx is 3.15. In this application, $\theta = \theta_a = 1/a = .1$, and x corresponds to c_1. Consequently, we have:

$$.1c_1 = 3.15 \rightarrow c_1 = 31.5 .$$

Therefore, the size .05 test based on $\hat{\lambda}_5$ would reject H_0 if $\hat{\lambda}_5 > 31.5$.

Now consider the test based on $\sum_1^k X_i$. In this case, c_2 is chosen so that:

$$\sup_{\Theta_0} P\left\{\sum_1^k X_i > c_2\right\} = \alpha .$$

Once again, the rejection probability is maximized over Θ_0 when $\lambda_1 = \cdots = \lambda_k = a$. For this value of $\underline{\lambda}$, $\sum_1^k X_i$ has a gamma distribution with parameters (k, a). Consequently, c_2 is seen to be the $(1 - \alpha)$th percentile of a $G(k, a)$ distribution.

For the third test, based on X_k, it is easily seen that c_3 should be chosen so that $P_{\lambda_k=a}\{X_k > c_3\} = \alpha$. We, therefore, obtain:

$$e^{-c_3/a} = \alpha \qquad \text{or} \qquad c_3 = -a(\ln \alpha) \ .$$

Unlike the first hypothesis testing situation considered, we do not have a consistent ordering of power functions due to the fact that the tests no longer have a critical value in common. Figure 10 gives several examples of how the 3 tests behave when testing $H_0 : \lambda_3 \leq 10$ vs. $H_1 : \lambda_3 > 10$ when $\alpha = .05$ and $k = 3$. In Figure 10(a) H_0 is true and use of $\bar{X}$ is seen to give the highest probability of acceptance. Unfortunately, this high acceptance probability carries over to cases where H_0 is false. Figures 10(b) and 10(c) show situations where the test based on $\bar{X}$ is biased. This is obviously a very undesirable trait. The tests based on $\hat{\lambda}_3$ and X_3 are fairly close in power. X_3 tends to perform better for values in Θ_1 which are fairly "close" to Θ_0, while $\hat{\lambda}_3$ does better for values in Θ_1 which are far from Θ_0. Again, the choice is up to the user, but tests based on $\bar{X}$ are not recommended due to bias.

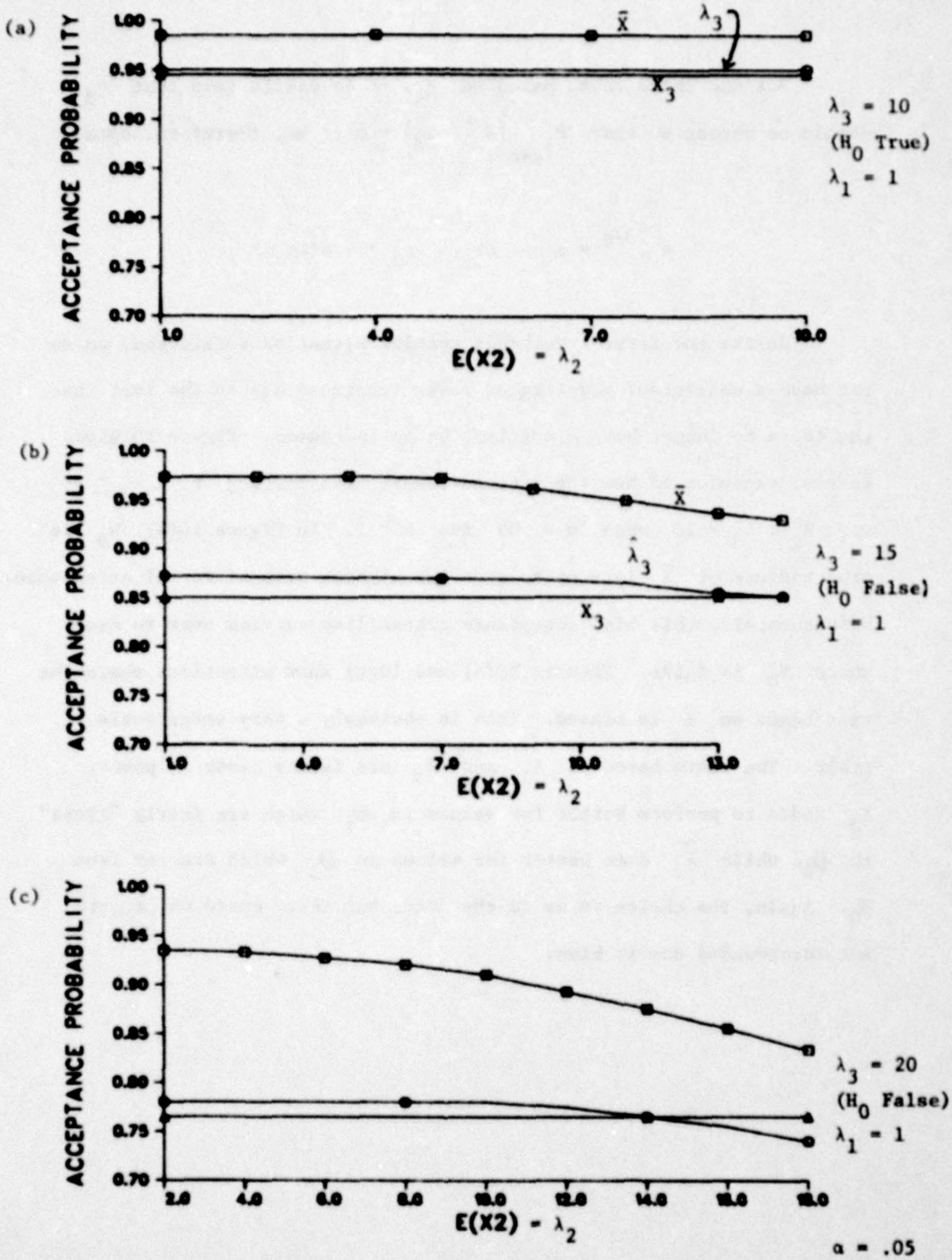
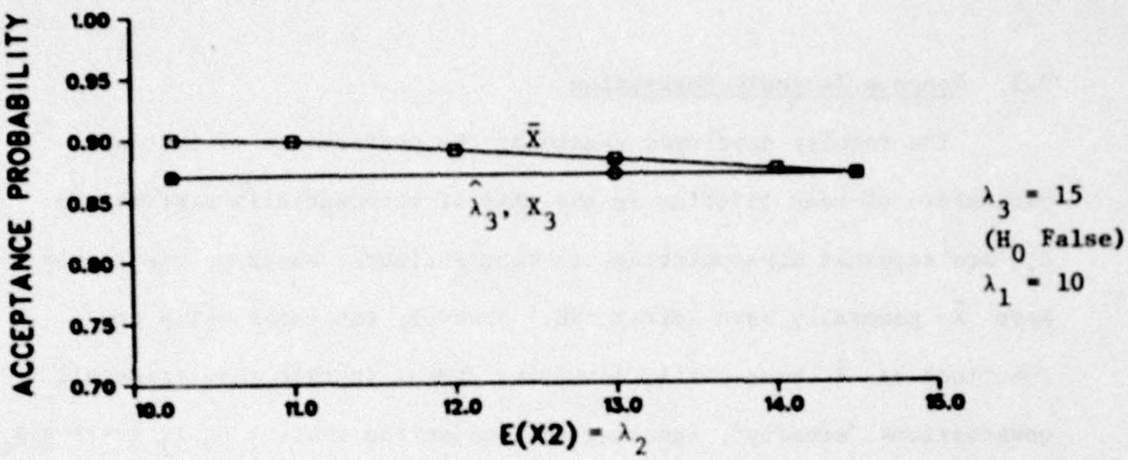
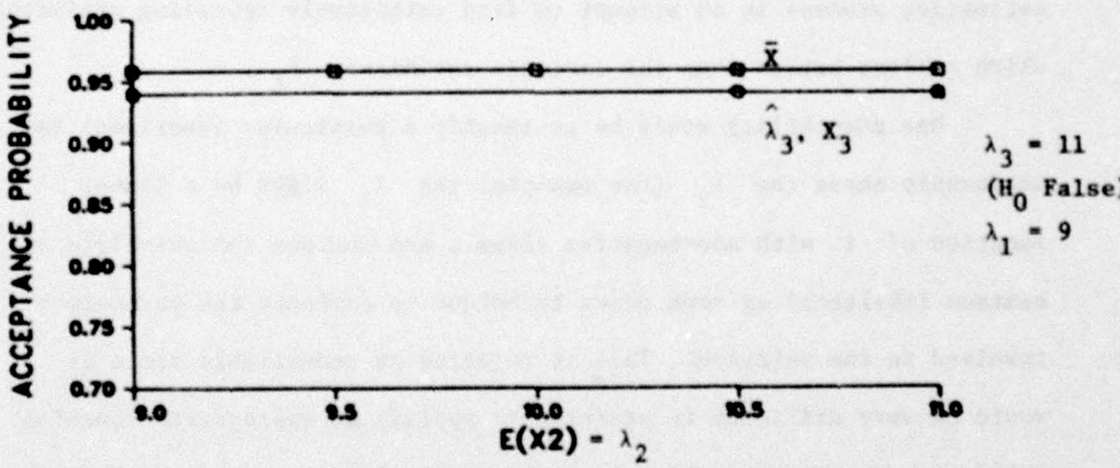


Figure 10. Comparative Behavior of Tests for H_0: $\lambda_3 \leq 10$

(d)



(e)



α = .05

Figure 10 - Continued

CHAPTER III

CONCAVE ISOTONIC ESTIMATORS

3.1. Concave Isotonic Regression

The results developed regarding the performance of isotonic estimators of mean lifetime in the case of exponentially distributed X_i are somewhat disappointing, in that estimates based on the sample mean $\bar{X}$ generally have better MSE. However, estimates which are functions of $\bar{X}$ have little intuitive appeal in that they treat all observations "equally", ignoring the assumption that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$. For example, early observations are given just as much weight as more recent ones in estimating λ_k, the final mean. This is somewhat counter-intuitive, and consequently we wish to modify the original model and estimation process in an attempt to find intuitively appealing estimators which perform better than the isotonic estimators $\hat{\lambda}_i$.

One possibility would be to specify a particular functional relationship among the λ_i (for example, the λ_i might be a linear function of i with non-negative slope), and then use the principle of maximum likelihood or some other technique to estimate the parameters involved in the relation. This is rejected as undesirable since it would be very difficult in practice to specify an appropriate function unless one knew quite a bit about the particular prototype development process involved. However, a reasonable modification to the original model might be the assumption that, in addition to an improvement in mean lifetime at each stage, the "big" improvements occur early in the

process as major design flaws are discovered and corrected while later improvements are limited to minor "fine tuning" changes. This could be modeled by assuming that the λ_i are non-decreasing, as before, but with the additional provision that the changes $(\lambda_{i+1} - \lambda_i)$ are decreasing. This is equivalent to saying that the λ_i are a concave non-decreasing function of i.

The proposed method of estimation, using the modified model described above, is to parallel the isotonic regression concept by finding values $\lambda_1^*, \lambda_2^*, \ldots, \lambda_k^*$ which minimize $\sum_{i=1}^{k} (Y_i - X_i)^2$. In this case the minimum is taken over all values $Y_1, \ldots, Y_k$ where

$$Y_1 \leq Y_2 \leq \cdots \leq Y_k$$

and $Y_{i+2} - Y_{i+1} \leq Y_{i+1} - Y_i$ for $i = 1, 2, \ldots, k - 2$.

Recall that in the case of regular isotonic regression and exponentially distributed X_i, minimizing the sum of squared deviations as given above was seen to be equivalent to maximizing the likelihood function. This equivalence does not hold in the "concave isotonic" case, so we are no longer dealing with maximum likelihood estimates. However, the concept of least-squares estimation is a traditional one, and would seem to be especially appropriate when dealing with mean-square-error.

The computation of concave isotonic estimates is considerably more complex than the regular isotonic estimates. First consider the simplest case, when $k = 3$. In this case, the values for the estimates

are easily found once the configuration of the X_i is known. Table 3 gives formulas for the λ_i^* according to which of seven regions in 3-space the vector (X_1, X_2, X_3) falls into.

If k is larger than 3, the computations increase in complexity and exact formulas for the λ_i^* are not found. However, the minimization problem involved can be expressed in a familiar form and the solution determined using any of a variety of algorithms. First note that:

$$\sum_{i=1}^{k} (Y_i - X_i)^2 = \sum_{i=1}^{k} Y_i^2 - 2 \sum_{i=1}^{k} X_i Y_i + \sum_{i=1}^{k} X_i^2 .$$

Since we wish to minimize this expression over values of the Y_i the final term, involving only X's, need not be considered. Also, multiplication of the entire function by the positive constant 1/2 will have no effect on the solution vector. Therefore, writing the function and constraints in matrix form, the problem can be expressed as follows:

$$\text{Minimize: } (1/2)\ \underline{y}^T \underline{By} + \underline{c}^T \underline{y}$$
$$\text{subject to: } \underline{Ay} \leq \underline{0}$$

where

$$\underline{B} = \underline{I}_k \qquad (k \times k \text{ identity matrix})$$

$$\underline{c} = -\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} \qquad \underline{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{pmatrix}$$

$$\underline{A} = \begin{bmatrix} 1 & -1 & 0 & \cdots & & & & 0 \\ 0 & 1 & -1 & \cdots & & & & 0 \\ \vdots & & 1 & -1 & & & & \vdots \\ 0 & 0 & \cdots & & & & 1 & -1 \\ 1 & -2 & 1 & 0 & \cdots & & & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ & & \vdots & & & & & \\ 0 & \cdots & & & 0 & 1 & -2 & 1 \end{bmatrix} \quad \begin{array}{l} (2k-3) \times k \\ \text{constraint matrix} \end{array}$$

When expressed in this form, the minimization problem is seen to be a classic quadratic programming problem. Consequently, any quadratic or non-linear programming algorithm can be used to find the λ_i^*.

As in the regular isotonic regression case, estimator performance has been studied using Monte Carlo techniques. Results of this study are given in Section 3.3.

3.2. An Alternative Concave Regression Algorithm

One difficulty with the concave isotonic approach discussed in the previous section is the complexity of the calculations involved. A potential user may not have a quadratic programming algorithm readily available and, due to difficulties encountered in learning or programming one, may choose to discard a potentially valuable technique. Therefore, an alternative method for generating a set of concave estimates is

suggested. The resulting estimates do not always minimize the restricted sum of squared deviations from the original observation (X_i), but they do compare favorably with the concave isotonic estimates and are much easier to obtain computationally. The steps involved are as follows:

(1) Find the isotonic regression, say $X_1^1, X_2^1, \ldots, X_k^1$, of the original data $X_1, X_2, \ldots, X_k$. Any of the techniques given in Barlow, et. al. could be used here. This will yield a set of estimates which will be non-decreasing but which may not be concave.

(2) Compute the differences $d_i = X_{i+1}^1 - X_i^1$ for $i = 1, 2, \ldots, k - 1$. We wish to modify the d_i to insure that they are non-increasing (but remain positive).

(3) Find the antitonic regression $d_1^1, \ldots, d_{k-1}^1$ of $d_1, \ldots, d_k$. The antitonic regression is defined to be the set of values $d_1^1, \ldots, d_{k-1}^1$ which minimize $\sum_{i=1}^{k-1} (Y_i - d_i)^2$ subject to the constraint $Y_1 \geq Y_2 \geq \cdots \geq Y_{k-1}$. Algorithms for finding the isotonic regression generally have an analogous counterpart for finding the antitonic regression.

(4) Use the differences d_i^1 to generate a set of estimates. A starting point, say a, is required in order to do this. Choose a to minimize the sum of squared deviations of the estimates from the original data values. The formula

for a is found as follows:

Let $\alpha_1 = d_1^1$

$$\alpha_2 = d_1^1 + d_2^1$$

$$\vdots$$

$$\alpha_{k-1} = d_1^1 + d_2^1 + \cdots + d_{k-1}^1 .$$

The estimates are found by letting

$$\lambda_1^1 = a + \alpha_0 \qquad \text{(define } \alpha_0 = 0\text{)}$$

$$\lambda_2^1 = a + \alpha_1$$

$$\vdots$$

$$\lambda_k^1 = a + \alpha_{k-1} \qquad (3.2.1)$$

where a is chosen to minimize $\sum_{i=1}^{k} (\lambda_i^1 - X_i)^2$ or $\sum_{i=1}^{k} (a + \alpha_{i-1} - X_i)^2$. Differentiating with respect to a and setting the derivative equal to zero, we obtain the formula:

$$a = \frac{\sum_{i=1}^{k} X_i - \sum_{i=1}^{k-1} \alpha_i}{k}$$

Once a is known, the λ_i^1 are found using formula (3.2.1). Experience shows that the λ_i^1 as calculated above will in fact be equal to the λ_i^* (concave isotonic regression estimates) in many

cases. In fact, considering the case $k = 3$ and referring back to Table 3, it can be shown that the two procedures give equivalent results except in some instances of case V.

As an example of the application of this algorithm, suppose that $k = 5$ and the observed lifetimes are 2, 8, 4, 12, 10. Isotonic and antitonic regression will be done using the "pool-adjacent-violators" algorithm given in Barlow, Bartholomew, Bremner, and Brunk, page 13 [1]. This algorithm simply pools any adjacent pair of observations (and their associated weights) which violate the desired ordering relation. This is done repeatedly until the ordering relation holds, starting in our case with unit weights. We begin by finding the isotonic regression of the X_i.

Lifetimes	2	8	4	12	10
Weights	1	1	1	1	1
1st Pooling	2	6	11		
Weights	1	2	2		

After one pooling we obtain for the isotonic regression the values 2, 6, 6, 11, 11. Consequently, the differences d_i are given by 4, 0, 5, 0. Since these are not non-increasing, we again pool-adjacent-violators, this time finding the antitonic regression.

Differences	4	0	5	0
Weights	1	1	1	1
1st Pooling	4	2.5	0	
Weights	1	2	1	

Once again the process terminates after only 1 pooling, and for the d_i^1 we obtain the values 4, 2.5, 2.5, 0. Now compute the α_i and a:

$$\alpha_0 = 0 \qquad \alpha_3 = 6.5 + 2.5 = 9$$

$$\alpha_1 = 4 \qquad \alpha_4 = 9 + 0 = 9$$

$$\alpha_2 = 4 + 2.5 = 6.5$$

$$a = \frac{(36 - 28.5)}{5} = 1.5 \; .$$

Using formula 3.2.1, we finally obtain the estimates:

$$\lambda_1^1 = 1.5 \quad \lambda_2^1 = 5.5 \quad \lambda_3^1 = 8.0 \quad \lambda_4^1 = \lambda_5^1 = 10.5 \; .$$

As in the concave isotonic regression case, estimator performance (in the exponential case) has been evaluated via Monte Carlo techniques. Results are given in the following section.

3.3. Performance of Concave Estimators (Exponential Case)

Computation of distribution functions and exact moments is even more difficult for the concave estimators than for the regular isotonic estimators. Consequently, evaluation of the performance of the λ_i^* and λ_i^1 extimators developed in Sections 3.1 and 3.2 has been limited to simulation studies, and no analytic results are available. However, the studies consistently show the superiority

of both types of concave estimators to the original isotonic estimators when the concavity assumption does indeed hold.

Figure 11 illustrates the results of a typical simulation study. In this case, the number of stages was set to 5 and, as in previous examples, the X_i were assumed to be exponentially distributed. It was assumed that the λ_i varied according to the arbitrary function $\lambda_i = A - B/i + B/k$ $(k=5)$. Thus, the value for A gives the final mean of the process, while B gives a measure of the "spread" or amount of improvement through the development process. Mean-square-errors are compared, using $\bar{\bar{X}}$ as a base, both for estimates of the final mean λ_k and for total mean-square-error over all λ_i.

As the graphs show, there is a consistent ordering of mean-square-errors over all values of B. The concave regression estimates (λ_i^*) show a very significant improvement over the original isotonic estimates $(\hat{\lambda}_i)$. When considering the estimation of λ_k alone, the mean-square-error for λ_k^* is consistently less than half that of $\hat{\lambda}_k$. The improvement in total mean-square-error is not quite as dramatic, but is also substantial.

The concave estimates of Section 3.2 (λ_i^1) consistently fall between the λ_i^* and the $\hat{\lambda}_i$. Although they do not perform as well as the λ_i^*, they are also superior to the $\hat{\lambda}_i$ while retaining a computational advantage over the λ_i^*.

It would appear that the concave estimators yield a substantial improvement over the isotonic estimators. Note, however, that all MSE ratios illustrated in Figure 16 are larger than one. Consequently, $\bar{\bar{X}}$

# of stages = k = 5 $\qquad \lambda_i = A - B/i + B/5 \qquad A = 10$

$\hat{\lambda}_i$ = isotonic estimate

λ_i^* = concave regression (Sec. 3.1)

λ_i^1 = alternate concave estimates (Sec. 3.2)

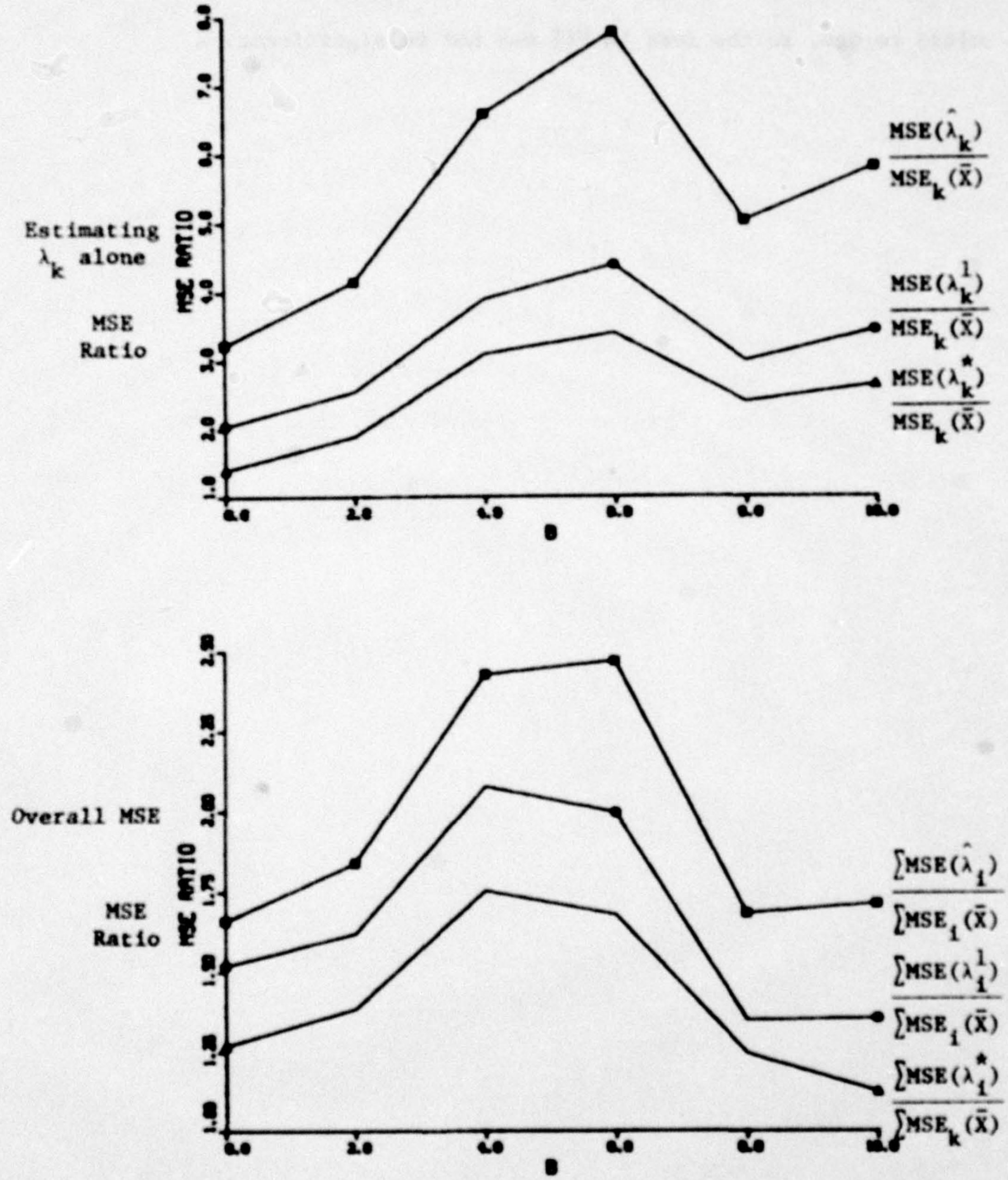


Figure 11. Performance of Concave Estimates

is still the superior estimator (in terms of MSE) of the λ_i. The choice of a concave estimator over $\bar{X}$ would have to be justified by balancing the loss in MSE with the concave estimator's greater intuitive appeal. Note that many of the MSE ratios are fairly close to one, so the loss in MSE may not be significant.

CHAPTER IV

ESTIMATOR PERFORMANCE WITH NON-EXPONENTIAL LIFETIMES AND RELIABILITY ESTIMATION

4.1. Isotonic Estimates Based on Multiple Samples per Stage

In previous sections, we have assumed that at each stage a single prototype with an exponentially distributed lifetime is tested. Let us continue the exponential assumption but suppose, in addition, that, at each development stage, several identical prototypes are built and tested. For simplicity assume that the same number of prototypes, say n, are built and tested at each stage. Let $X_{i1}, X_{i2}, \ldots, X_{in}$ represent the observed lifetimes at stage i. Then the joint density of all observations can be written:

$$f_{\underline{x}}(x_{11}, \ldots, x_{1n}, \ldots, x_{k1}, \ldots, x_{kn}) = \prod_{i=1}^{k} \prod_{j=1}^{n} \frac{1}{\lambda_i} e^{-x_{ij}/\lambda_i}$$

$$= \prod_{i=1}^{k} \frac{1}{\lambda_i^n} e^{-\frac{1}{\lambda_i}\left[\sum_{j=1}^{n} x_{ij}\right]} = \prod_{i=1}^{k} \frac{1}{\lambda_i^n} e^{-\frac{n}{\lambda_i}\bar{x}_i}$$

where $\bar{x}_i$ is the mean of the observed lifetimes at stage i.

Our goal, as usual, is to efficiently estimate the λ_i with particular emphasis on λ_k, the final mean. Since the joint density of all the observations is seen to depend on the observations only through the $\bar{X}_i$, we know from the factorization theorem (c.f. Ferguson, page 115 [10]), that $(\bar{X}_1, \ldots, \bar{X}_k)$ is jointly sufficient for $(\lambda_1, \ldots, \lambda_k)$. Consequently, we need only consider

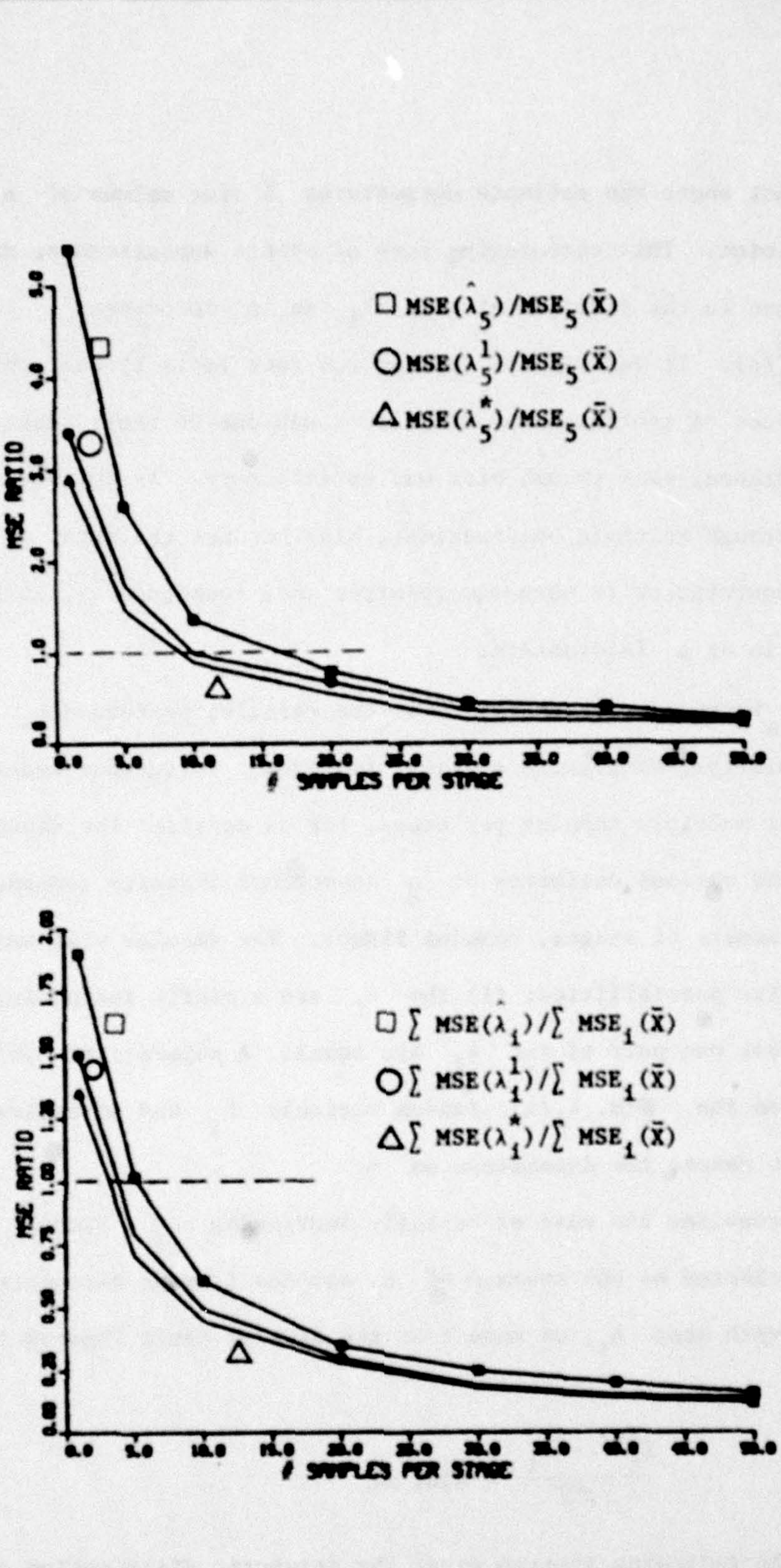
estimators which are functions of the $\bar{X}_i$. Since $\bar{X}_i$ is easily seen to have a gamma distribution with parameter vector $(n, \lambda_i/n)$, let us simplify matters by assuming that at each stage i a $\mathcal{G}(n, \lambda_i/n)$ random variable X_i is observed where n is known and $\lambda_1 \leq \cdots \leq \lambda_k$.

As in the previously examined exponential case, Barlow, Bartholomew, Bremner, and Brunk, page 99 (1972), show that restricted likelihood estimates for the λ_i are given by the isotonic regression of the X_i with unit weights (assuming an equal number of samples at each stage). We will consider the same estimation procedures discussed earlier and focus attention on the effect of multiple samples at each stage. No exact distribution theory has been developed in this case, as the mathematics involved is very awkward even in the simplist case when the λ_i are all equal. Consequently, all results given are due to simulation studies.

As an illustration, suppose that $k = 5$ stages and that the means are both increasing and concave, according to the formula $\lambda_i = A - B/i + B/k$. Letting $A = 10$ and $B = 8$, we obtain:

$$\lambda_1 = 3.60 \qquad \lambda_2 = 7.60 \qquad \lambda_3 = 8.93 \qquad \lambda_4 = 9.60 \qquad \lambda_5 = 10.00\ .$$

Figure 12 compares mean-square-errors for various estimation procedures as a function of n, the number of samples per stage. Note that as n increases, the performance of the isotonic and concave isotonic estimators steadily improves relative to that of $\bar{X}$, the overall mean. For all three types of isotonic estimators considered there is a



$k = 5 \qquad \lambda_i = 10 - 8/i + 8/5$

Figure 12. Comparing MSE to Sample Size

threshold point where the estimate outperforms $\bar{X}$ for values of n beyond that point. This encouraging turn of events appears to be due to the decrease in the variance of each X_i as n increases ($Var(X_i) = \lambda_i^2/n$). It was seen in Section 2.4 (see Table 1) that the poor performance of isotonic-type estimators was due to their inability to reduce variance, even though bias was satisfactory. As the variance is reduced through multiple observations, bias becomes the more significant contributor to mean-square-error and, consequently, it is now $\bar{X}$ that is at a disadvantage.

As n increases, it is seen that the relative performance of the isotonic-type estimators steadily improves. To further examine the effects of multiple samples per stage, let us consider the asymptotic behavior of the various estimates as n approaches infinity (assume that k, the number of stages, remains fixed). The results will vary according to two possibilities: (1) the λ_i are strictly increasing, and (2) at least one pair of the λ_i are equal. A superscript (n) will be used on the $\mathcal{G}(n, \lambda_i/n)$ random variable X_i and on estimates of the λ_i to denote the dependence on n.

First consider the case of strictly increasing λ_i. Since $X_i^{(n)}$ is distributed as the average of n samples from an exponential distribution with mean λ_i, we know from the Central Limit Theorem that

$$\frac{X_i^{(n)} - \lambda_i}{\lambda_i/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

as $n \to \infty$. The following theorem gives the asymtotic distribution of the isotonic estimates $\hat{\lambda}_i^{(n)}$.

<u>Theorem 4.1</u>: Assume $\lambda_1 < \lambda_2 < \cdots < \lambda_k$. Then for any i we have:

$$\lim_{n\to\infty} P\{\hat{\lambda}_i^{(n)} = X_i^{(n)}\} = 1 .$$

Consequently, $\hat{\lambda}_i^{(n)}$ and $X_i^{(n)}$ have the same asymptotic distribution:

$$\frac{\hat{\lambda}_i^{(n)} - \lambda_i}{\lambda_i/\sqrt{n}} \xrightarrow{d} N(0,\ 1) .$$

The theorem should be intuitively clear for the following reason. If the observed $X_i^{(n)}$ are properly ordered, then their isotonic regression is simply the $X_i^{(n)}$ themselves - no changes need to be made. Since the means are strictly increasing and the variances approach zero as $n \to \infty$, the $X_i^{(n)}$ are going to be properly ordered with probability approaching one. Consequently, the $\lambda_i^{(n)}$ will equal the $X_i^{(n)}$ with ever-increasing probability. A more rigorous proof follows:

<u>Proof</u>: Let $\epsilon = \max\limits_{2\le i\le k} (\lambda_i - \lambda_{i-1})$. Let A_n denote the event that the isotonic regression is equal to the original observations: $A_n = [\hat{\lambda}_i^{(n)} = X_i^{(n)}\ \forall_i]$. Note that A_n will definitely occur if each $X_i^{(n)}$ is within $\epsilon/2$ of λ_i. Now,

$$P(A_n) \geq P\{|X_i^{(n)} - \lambda_i| \leq \epsilon/2\ \forall_i\}$$

$$= \prod_{i=1}^{k} P\{|X_i^{(n)} - \lambda_i| \leq \epsilon/2\} \qquad \text{(independence)}$$

$$\geq \prod_{i=1}^{k} \left[1 - \frac{\mathrm{Var}(X_i^{(n)})}{\epsilon^2/4} \right] \qquad \text{(Chebychev inequality)}$$

$$= \prod_{i=1}^{k} \left[1 - \frac{4\lambda_i^2}{n\epsilon^2} \right]$$

$$\geq \prod_{i=1}^{k} \left[1 - \frac{4\lambda_k^2}{n\epsilon^2} \right] \qquad (\lambda_1 < \cdots < \lambda_k)$$

$$= \left[1 - \frac{4\lambda_k^2}{n\epsilon^2} \right]^k .$$

Letting $n \to \infty$, we obtain: $\lim_{n\to\infty} P(A_n) = 1$. Consequently, for any i $\lim_{n\to\infty} P\{\hat{\lambda}_i^{(n)} = X_i^{(n)}\} = 1$.

Now let

$$B_n = \frac{X_i^{(n)} - \lambda_i}{\lambda_i/\sqrt{n}} \quad \text{and} \quad C_n = \frac{\hat{\lambda}_i^{(n)} - \lambda_i}{\lambda_i/\sqrt{n}} .$$

Then, for any positive δ,

$$P\{|B_n - C_n| < \delta\} > P\{|B_n - C_n| = 0\} = P\{B_n = C_n\} .$$

So, $P\{|B_n - C_n| < \delta\} \to 1$ as $n \to \infty$, or $B_n - C_n \xrightarrow{P} 0$. Consequently, using Stutsky's Theorem (refer to the proof in Section 2.6), B_n and C_n have the same asymptotic distribution. □

Note that if the λ_i *are strictly increasing* *__and__* *strictly concave, an analogous proof could be constructed to show that the concave isotonic estimates* $(\lambda_i^{*(n)}$ *or* $\lambda_i^{1(n)})$ *also have the same*

asymptotic distributions as the $X_i^{(n)}$.

Now consider the second case, where at least one pair of the λ_i are equal. In this case we do not have $P\{\hat{\lambda}_i^{(n)} = X_i^{(n)}\}$ converging to one, and an exact limiting distribution is not obtained. However, in theorems 2.1 and 2.2 of Barlow, Bartholomew, Bremner, and Brunk [1], it is seen that if $\lambda_1 \leq \cdots \leq \lambda_k$ and $Y_1^{(n)}, \cdots Y_k^{(n)}$ are a consistent set of estimates of the λ_i then the isotonic regression $Y_i^{(n)*}$ of the $Y_i^{(n)}$ are also a consistent set of estimates. It is also seen that:

$$\sum (Y_i^{(n)*} - \lambda_i)^2 \leq \sum (Y_i^{(n)} - \lambda_i)^2 . \qquad (4.1.1)$$

In our case the $X_i^{(n)}$ may be thought of as consistent estimates of the λ_i, so we can claim that the $\hat{\lambda}_i^{(n)}$ are also consistent. Note that the above inequality (taking expectations on both sides) shows that total mean-square-error is always improved when the isotonic estimates are used rather than the individual observations.

The proofs presented by Barlow, Bartholomew, Bremner, and Brunk [1] to show inequality (4.1.1) and consistency of isotonic estimates can be paralleled to achieve the same results in the case of the concave isotonic estimates $\lambda_i^{*(n)}$. Consequently, it is claimed that the $\lambda_i^{*(n)}$ are also consistent estimates of the λ_i (assuming that the λ_i are indeed isotonic and concave) and that total mean-square-error is always reduced when using concave isotonic estimates rather than the original observations.

4.2. Estimation of Reliability in the Exponential Case

In many cases involving lifetime data the ultimate goal is to estimate the reliability at a given time t $(P\{X > t\})$ rather than the mean lifetime. For example, in the prototype development model we may be interested in the probability that the device will meet a predetermined lifetime specification. In the case of exponentially distributed lifetimes, the reliability at time t is given by $R(t) = e^{-t/\lambda}$ where λ represents the mean lifetime as before. To proceed in the same manner as when considering the mean lifetime estimation problem, we would like to develop efficient estimators of the parameters $R_1(t), R_2(t), \ldots, R_k(t)$ where $R_i(t) = e^{-t/\lambda_i}$. The assumption that $R_1(t) \leq R_2(t) \leq \cdots \leq R_k(t)$ will be made, analogous to the previous assumption of non-decreasing lifetimes. Note that the time t is assumed to be known and fixed.

Let us begin by finding the restricted maximum likelihood estimators of the $R_i(t)$. Let

$$A = \{\underline{\lambda} \mid \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k\}$$

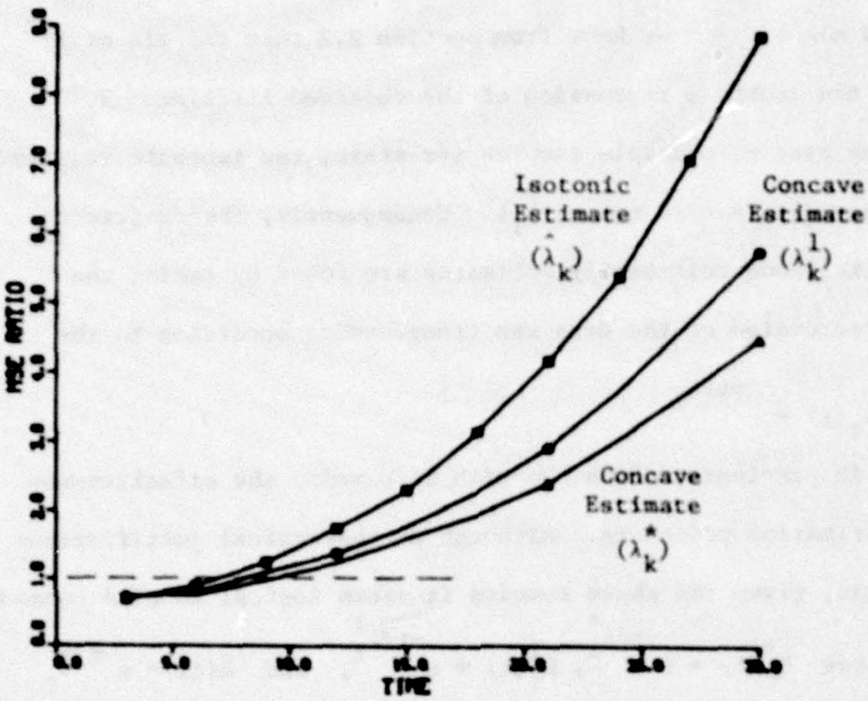
$$B = \{\underline{R(t)} \mid R_1(t) \leq R_2(t) \leq \cdots \leq R_k(t)\},$$

and define $f : A \to B$ by $f(\lambda_1, \ldots, \lambda_k) = \left(e^{-t/\lambda_1}, \ldots, e^{-t/\lambda_k}\right) = \underline{R(t)}$. The restricted maximum likelihood estimators for the $R_i(t)$ are found by maximizing the likelihood function (reparameterized in terms of the $R_i(t)$) over the set B. However, since f is a 1-1 function from

A to B, it is seen by the maximum likelihood invariance property (Mood, Graybill, and Boes, page 285 [15]) that the restricted mle's of the $R_i(t)$ (or $f(\underline{\lambda})$) are given by $f(\hat{\underline{\lambda}})$, where $\hat{\underline{\lambda}}$ represents the restricted mle of $\underline{\lambda}$. We know from Section 2.2 that the mle of $\underline{\lambda}$ is simply the isotonic regression of the observed lifetimes X_i (or, in the case of multiple samples per stage, the isotonic regression of the stage-wise sample means $\bar{X}_i$). Consequently, the restricted maximum likelihood reliability estimates are found by taking the isotonic regression of the data and transforming according to the formula $\hat{R}_i(t) = e^{-t/\hat{\lambda}_i}$.

As in previous studies, we wish to examine the effectiveness of this estimation procedure. Although no theoretical justification is presented, given the above results it seems logical to also consider the estimates $R_i^*(t) = e^{-t/\lambda_i^*}$, $R_i^1(t) = e^{-t/\lambda_i^1}$, and $\bar{R}(t) = e^{-t/\bar{X}}$, where λ_i^*, λ_i^1, and $\bar{X}$ are the previously defined competitive estimators of the λ_i. This study will be limited to considering the estimation of $R_k(t)$, the final or current reliability in the prototype development process.

Figures 13 and 14 give sample results of simulation studies made to evaluate estimator effectiveness. As in previous examples, mean-square-error is used as the evaluation criteria, and the results are given in terms of MSE ratios with the MSE of the estimator $\bar{R}(t) = e^{-t/\bar{X}}$ in the denominator. Figure 13 gives results when one sample is tested at each stage, while Figure 14 shows the effect of multiple samples per stage.



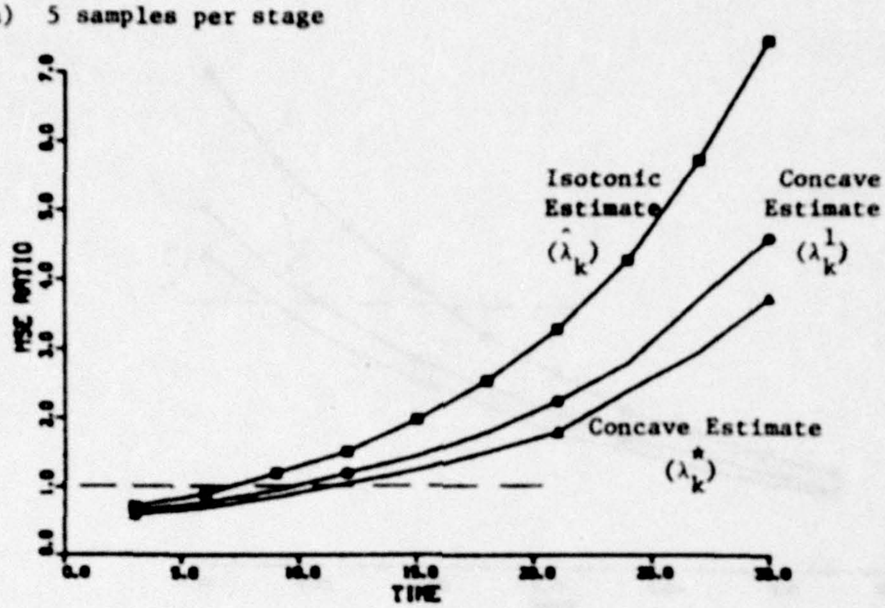
$k = 5$ Ratio denominator is $MSE_k(e^{-t/\bar{X}})$

$\lambda_i = 10 - 8/i + 8/5$

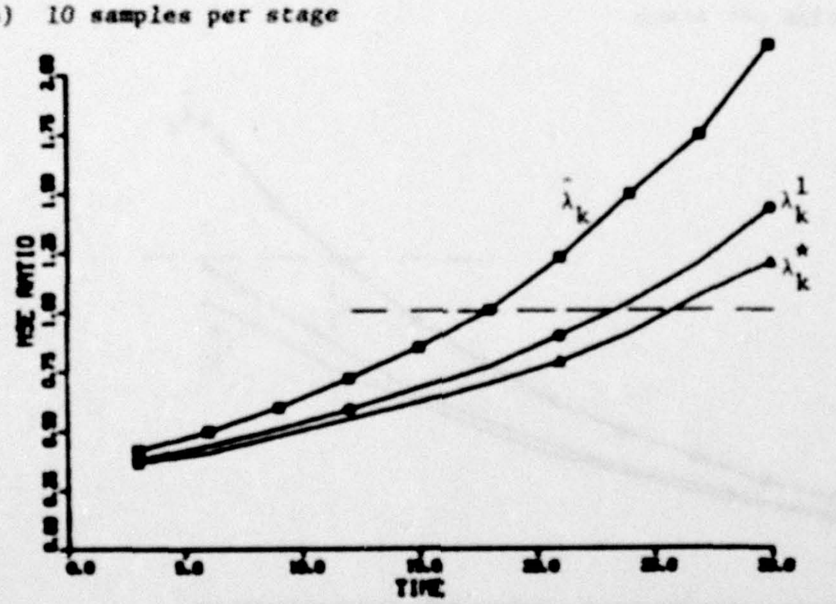
Samples per Stage = 1

Figure 13. Relative MSE When Estimating Reliability

(a) 5 samples per stage



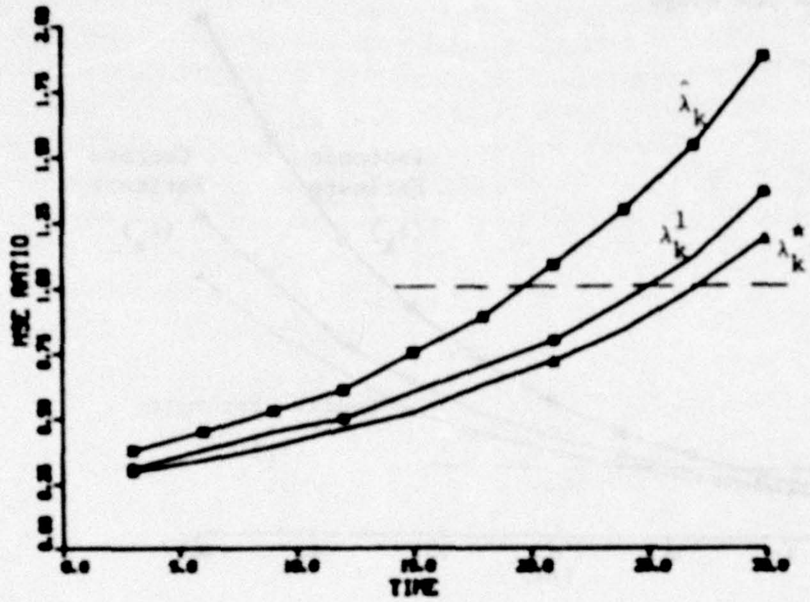
(b) 10 samples per stage



$k = 5 \quad \lambda_i = 10 - 8/i + 8/5 \qquad$ Ratio Denominator is $MSE_k(e^{-t/\bar{X}})$

Figure 14. Relative MSE When Estimating Reliability -Effect of Multiple Samples per Stage

(c) 15 samples per stage



(d) 20 samples per stage

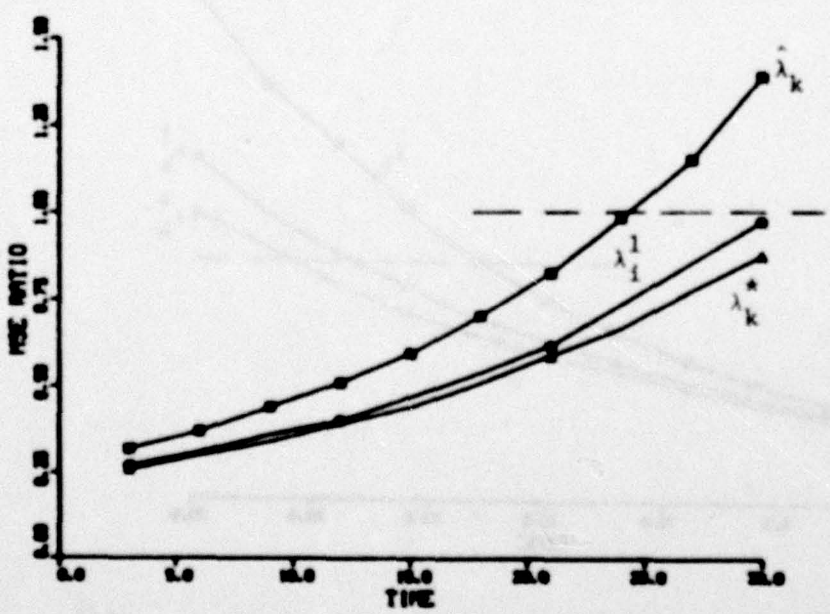


Figure 14. (continued)

Note that in all figures the previously encountered ordering of mean-square-errors among the three isotonic-based estimators is preserved during the transformation from mean lifetime estimation to reliability estimation. However, a new factor comes into play here - the time t at which the reliability is being estimated. In all cases it is seen that the isotonic-based estimators perform better relative to $e^{-t/\bar{X}}$ at earlier times. Table 4 gives a typical breakdown of MSE into its components of variance and bias at various times, showing the effect of the transformation to reliability estimation on both factors. Note in Figure 13 that if t is sufficiently small then the isotonic-based estimates are superior to $e^{-t/\bar{X}}$. As the number of samples per stage increases (see Figure 14) and a subsequent reduction in variance occurs, the range of t-values for which the isotonic estimators are superior to $e^{-t/\bar{X}}$ grows steadily. This is in agreement with earlier results (Section 4.1) concerning the increased effectiveness of isotonic-based estimators when variance is reduced.

As the number of samples per stage becomes large, asymptotic results similar to those of Section 4.1 can be stated. Letting $X_i^{(n)}$ represent a $\mathscr{G}\left(n, \frac{\lambda_i}{n}\right)$ random variable (corresponding to the mean of n samples at stage i), it is seen that $X_i^{(n)}$ is the unrestricted mle of λ_i and consequently by invariance $e^{-t/X_i^{(n)}}$ is the mle of $R_i(t) = e^{-t/\lambda_i}$. We know from maximum likelihood theory (Mood, Graybill, and Boes, page 358 [15]) that $e^{-t/X_i^{(n)}}$ is asymptotically normally

distributed and efficient. Now it was shown in Section 4.1 that if the λ_i are strictly increasing then $\lim_{n\to\infty} P(\hat{\lambda}_i^{(n)} = X_i^{(n)}) = 1$.

Consequently, $\lim_{n\to\infty} P\left\{ e^{-t/\hat{\lambda}_i^{(n)}} = e^{-t/X_i^{(n)}} \right\} = 1$ for any t and the isotonic reliability estimators are seen to have the same asymptotically normal distribution as the unrestricted maximum likelihood estimators. In the case of possible equal means, only consistency was shown in Section 4.1, and this consistency will still hold in the case of reliability estimation due to the continuity of the function $f(x) = e^{-t/x}$.

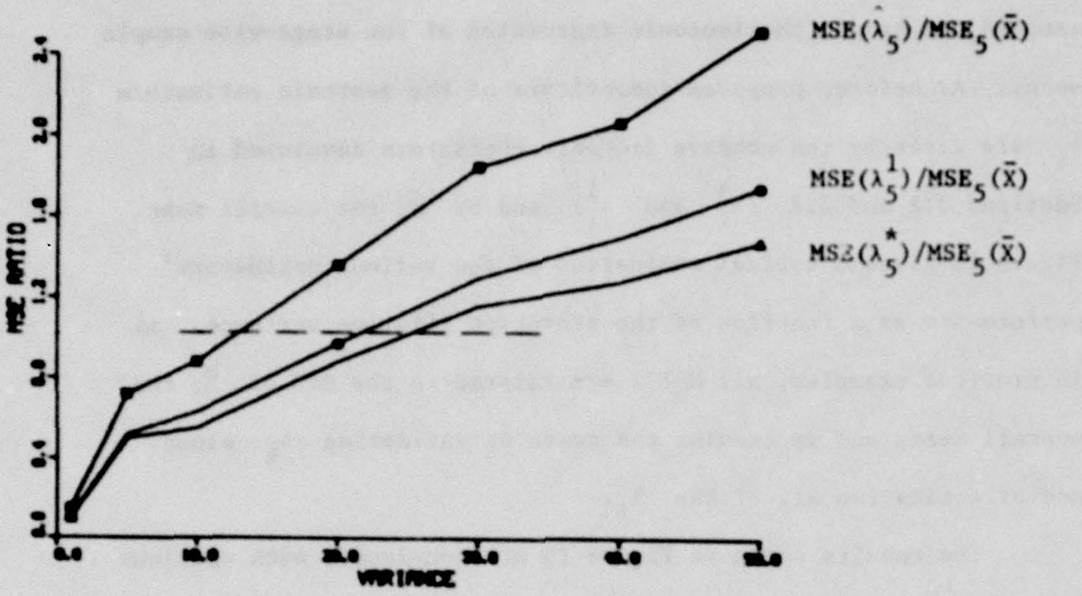
4.3. Estimation for Normal and Weibull Distributed Lifetimes

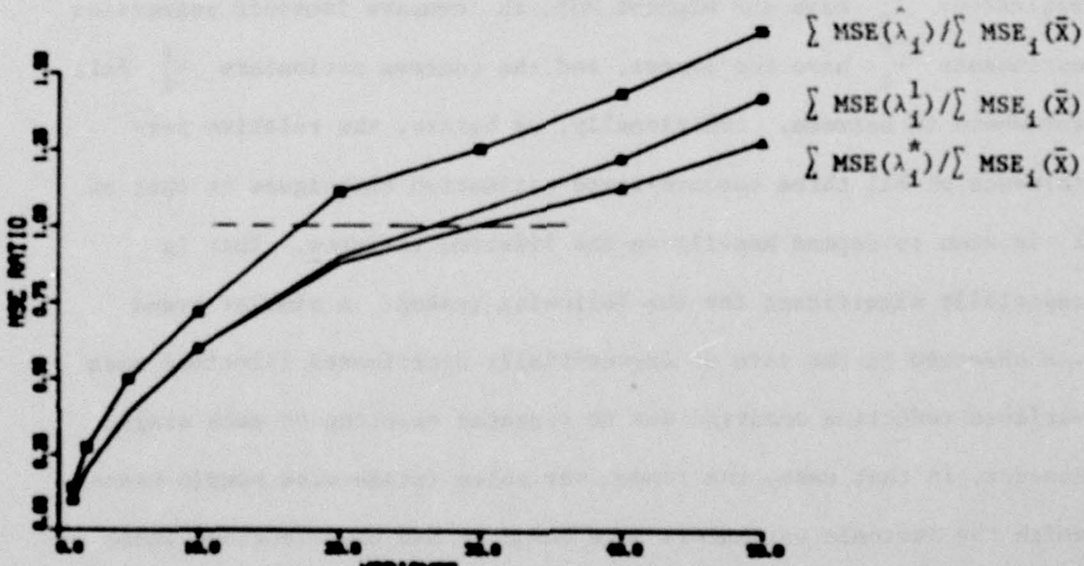
In previous studies we have assumed that we were dealing with exponentially distributed component lifetimes, yielding estimators based on either exponential random variables (one sample per stage) or gamma random variables (multiple samples per stage). We now wish to determine, using simulation techniques, whether the behavior of the istonic-based estimators changes radically when other common lifetime distributions are assumed.

Suppose that the prototype lifetimes are approximately normally distributed, where for simplicity we will assume that the variances at each stage are equal (but unknown). This is another situation (recall Section 1.2) where the restricted maximum likelihood estimators of the mean lifetimes λ_i are given by the isotonic regression of the observed lifetimes or, in the case of multiple

samples per stage, the isotonic regression of the stage-wise sample means. As before, proposed competitors of the isotonic estimators $\hat{\lambda}_i$ are given by the concave isotonic estimators developed in Sections 3.1 and 3.2 (λ_i^* and λ_i^1) and by $\bar{X}$, the overall mean. Figure 15 gives a typical evaluation of the various estimators' performance as a function of the prototype lifetime variance. As in previous examples, all MSE's are related to the MSE of $\bar{X}$, the overall mean, and we examine the cases of estimating λ_k alone and of estimating all of the λ_i.

The results shown in Figure 15 are consistent with previous studies in two important ways. First, the ordering of mean-square-errors among the isotonic-based estimators is consistent: the isotonic estimators $\hat{\lambda}_i$ have the highest MSE, the concave isotonic regression estimators λ_i^* have the lowest, and the concave estimators λ_i^1 fall somewhere in between. Additionally, as before, the relative performance of all three concave-based estimation techniques to that of $\bar{X}$ is seen to depend heavily on the lifetime variance. This is especially significant for the following reason. A similar trend was observed in the case of exponentially distributed lifetimes when variance reduction occurred due to repeated sampling at each stage. However, in that case, the random variables (stage-wise sample means) which the isotonic estimators were based on had distributions whose variance and shape changed as n, the number of samples per stage, increased. In the case of normally distributed lifetimes the shape is unchanged during variance reduction, and consequently, we can





$\lambda_i = A - B/i + B/k$ $\quad A = 10 \quad B = 8 \quad k = 5$

Figure 15. Evaluation of Estimates Based on Normally Distributed Lifetimes

safely attribute the increased effectiveness of the isotonic estimators to that reduction. Note that for the normal case the effect of taking multiple samples per stage is not shown, since this is seen to be equivalent to having one sample per stage with an appropriately reduced variance.

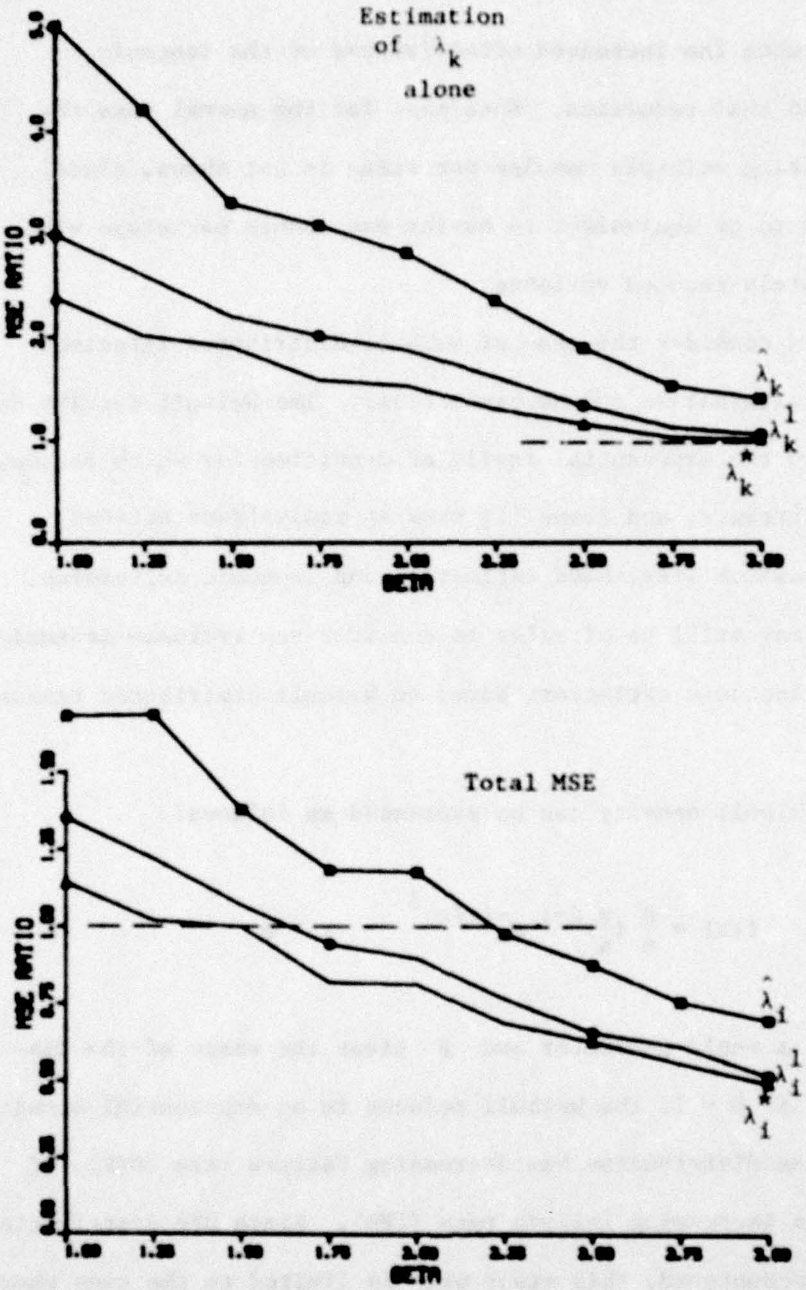
Let us consider the case of Weibull distributed lifetimes as a second alternative to the exponential. The Weibull density does not fall into the exponential family of densities for which Barlow, Batholomew, Bremner, and Brunk [1] show an equivalence between restricted maximum likelihood estimation and isotonic regression. However, it may still be of value to consider and evaluate isotonic and concave isotonic estimators based on Weibull distributed random variables.

The Weibull density can be expressed as follows:

$$f(x) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} e^{-(x/\eta)^{\beta}} \qquad x > 0$$

where η is a scale parameter and β gives the shape of the distribution. If $\beta = 1$, the Weibull reduces to an exponential density. If $\beta < 1$, the distribution has decreasing failure rate (DFR), if $\beta > 1$ it has increasing failure rate (IFR). Since DFR distributions are rarely encountered, this study will be limited to the case where $\beta > 1$.

Figure 16 gives a typical comparison of MSE's for the various estimators, showing the effect of varying the shape parameter β.



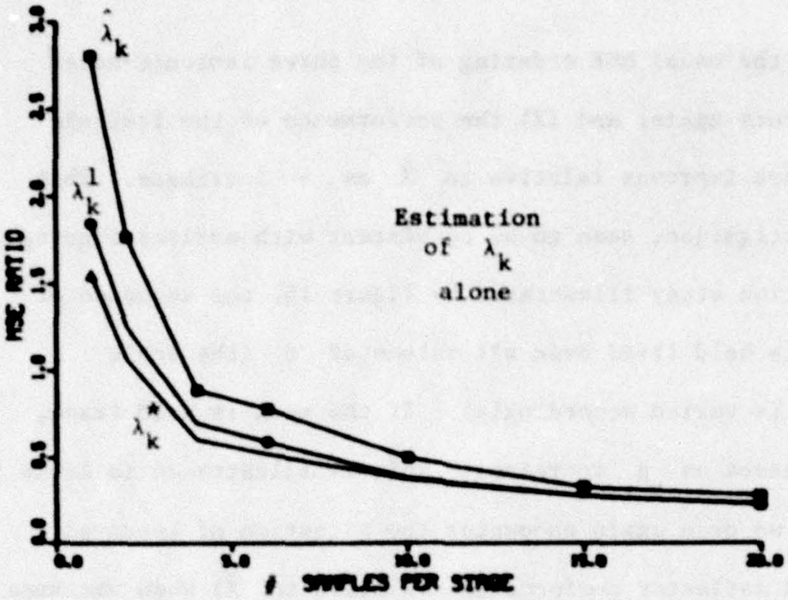
$k = 5 \qquad \lambda_i = 10 - 8/i + 8/5$

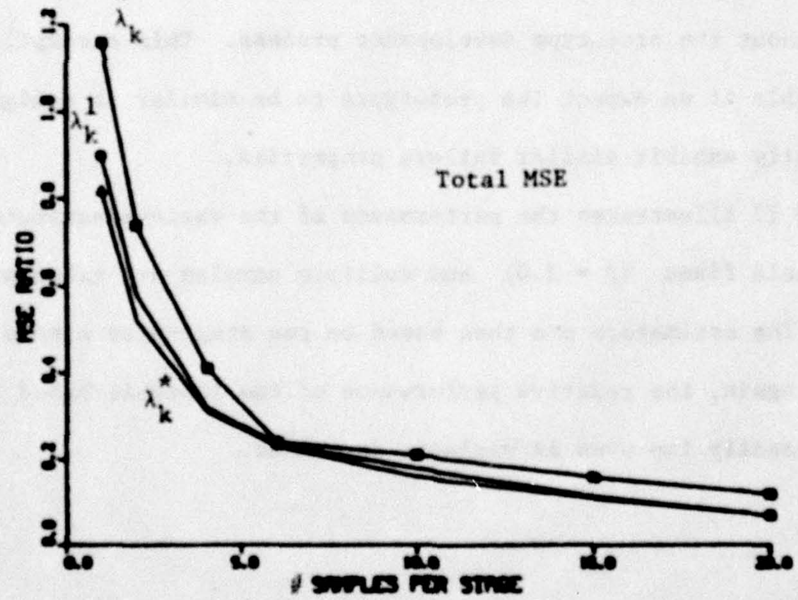
Figure 16. Evaluation of Estimates Based on Weibull-Distributed Lifetimes (Effect of varying shape parameter β)

Note that (1) the usual MSE ordering of the three isotonic-based estimators occurs again; and (2) the performance of the isotonic based estimators improves relative to $\bar{X}$ as β increases. This is, upon investigation, seen to be consistent with earlier findings. In the simulation study illustrated by Figure 16, the sequence of means $\{\lambda_i\}$ is held fixed over all values of β (the scale parameter, η, is varied accordingly). If the mean is held fixed, variance decreases as β increases. This is illustrated in Table 5. Consequently, we once again encounter the situation of improved isotonic-based estimator performance (relative to $\bar{X}$) when variance is reduced.

It is assumed in the Weibull simulations that β does not change throughout the prototype development process. This assumption seems reasonable if we expect the prototypes to be similar in design and consequently exhibit similar failure properties.

Figure 17 illustrates the performance of the various estimators when β is held fixed ($\beta = 2.0$) and multiple samples are taken at each stage. The estimators are then based on the stage-wise sample means. Once again, the relative performance of the isotonic-based estimators steadily improves as variance decreases.





$k = 5 \qquad \beta = 2.00 \qquad \lambda_i = 10 - 8/i + 8/5$

Figure 17. Evaluation of Estimators Based on Weibull-Distributed Lifetimes (Effect of varying samples per stage)

CHAPTER V

CONCLUSIONS

The purpose of this paper was to propose and evaluate various techniques for estimating lifetimes or reliabilities in a particular reliability growth or prototype development model. Evaluation involved several criteria, including accuracy (mean-square-error), large sample properties, usefulness in hypothesis testing situations, and intuitive appeal. The simple growth or prototype development model and the maximum likelihood principle led to the technique of isotonic regression, which in turn led to similar techniques involving the additional assumption of concavity. In addition, an estimate $(\bar{X})$ was evaluated which intuitively ignored the growth assumption and would perhaps be more appropriate in the classical case of identically distributed samples.

Due to computational difficulties encountered, exact results and analytical proofs were primarily limited to the simplest cases, particularly those dealing with exponentially distributed prototype lifetimes. However, simulation studies involving more complex situations lead one to believe that the general nature of the results are true for a fairly large class of common lifetime distributions.

The effectiveness of the isotonic-based estimators varies, as one might intuitively suspect, according to several factors. If there is little significant improvement from stage to stage in the

prototype development process, then we are essentially in the classical situation of independent identically distributed random variables and there is no reason to believe an isotonizing process will improve matters, which proves to be true. In addition, the isotonic estimators do relatively little to reduce variance, and consequently the overall mean may still prove superior (in terms of MSE, not intuitive appeal) in cases where there is a definite improvement in mean lifetime but this improvement is hidden by large sample variances. The situations where a significant advantage is gained through some sort of isotonizing process tend to be those where an improvement does exist throughout the process and the variances of the observed random variables are "moderate". Obviously, on the other end of the scale, the isotonic estimators show little improvement over the use of the observations themselves when variances are very small and the means are strictly increasing. This is due to the fact that the isotonizing process is essentially a rearrangement of the observed values into the "correct" or hypothesized ordering, and if the variances are sufficiently small the observations will be correctly ordered to begin with so no changes will take place. However, Barlow, Bartholomew, Bremner, and Brunk [1], do show that if the order assumptions (and this conclusion extends to concavity assumptions) are correct, then in total the isotonized values will be at least as good as the originals with respect to mean-square-error. Consequently, even with very small variances nothing but computational time will be lost, and some gain may occur. Note that a gain is not guaranteed when considering the estimation of any individual parameter.

AD-A075 578   STANFORD UNIV CA DEPT OF OPERATIONS RESEARCH   F/G 12/1
PROPERTIES OF ISOTONIC ESTIMATORS OF MEAN LIFETIME IN A SIMPLE --ETC(U)
AUG 79   T P MCWILLIAMS   N00014-75-C-0561
UNCLASSIFIED   TR-194   NL

2 OF 2
AD-
A075578

END
DATE
FILMED
11-79
DDC

The only large sample case investigation involving a large number of stages, rather than samples per stage, was restricted to the investigation of the final isotonic estimate and the case of exponentially distributed lifetimes. In this case exact results were obtainable, but were somewhat disappointing in that the usual convergence one would hope to see with increasing amounts of relevant data does not occur. The prototype development process may be converging to an optimal design with a corresponding optimal mean lifetime, but the isotonic estimate of the final mean in the process will not converge to this optimal value as its variance does not go to zero. This compares unfavorable with $\bar{X}$.

The extension from the regular isotonic estimators to concave isotonic estimators offers a significant improvement in performance at little expense in the generality of the assumptions. In particular, variance reduction is much better near the final stage, and the estimators have consistently proved superior to the isotonic ones in a wide variety of situations. The concave estimators proposed in Section 3.2 are almost as good as the concave regression estimators and have a significant computational advantage.

When considering the estimation of reliabilities rather than mean lifetimes, there is no real change in the conclusions regarding situations when the isotonic-based estimators do or do not perform well.

APPENDIX A

Distributions functions and expected values of $\hat{\lambda}_k$ for $k = 3$ and $k = 4$. Based on exponentially distributed X_i.

$$F_3(x) = 1 - e^{-\theta_3 x} + \frac{\theta_3}{\theta_3 - \theta_2}\left(e^{-(\theta_2+\theta_3)x} - e^{-2\theta_2 x}\right)$$

$$+ \frac{\theta_2\theta_3}{(\theta_2-\theta_1)(\theta_3-\theta_1)}\left(e^{-(2\theta_1+\theta_3)x} - e^{-3\theta_1 x}\right)$$

$$+ \frac{\theta_2\theta_3}{(\theta_2-\theta_1)(\theta_3-\theta_2)}\left(e^{-(\theta_1+2\theta_2)x} - e^{-(\theta_1+\theta_2+\theta_3)x}\right)$$

$$F_4(x) = 1 - e^{-\theta_4 x} + \frac{\theta_4}{\theta_4 - \theta_3}\left(e^{-(\theta_3+\theta_4)x} - e^{-2\theta_3 x}\right)$$

$$+ \frac{\theta_3\theta_4}{(\theta_3-\theta_2)(\theta_4-\theta_2)}\left(e^{-(2\theta_2+\theta_4)x} - e^{-3\theta_2 x}\right)$$

$$+ \frac{\theta_3\theta_4}{(\theta_3-\theta_2)(\theta_4-\theta_3)}\left(e^{-(\theta_2+2\theta_3)x} - e^{-(\theta_2+\theta_3+\theta_4)x}\right)$$

$$+ \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_1)(\theta_4-\theta_1)}\left(e^{-(3\theta_1+\theta_4)x} - e^{-4\theta_1 x}\right)$$

$$+ \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_1)(\theta_4-\theta_3)}\left(e^{-(2\theta_1+2\theta_3)x} - e^{-(2\theta_1+\theta_3+\theta_4)x}\right)$$

$$+ \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_2)(\theta_4-\theta_2)}\left(e^{-(\theta_1+3\theta_2)x} - e^{-(\theta_1+2\theta_2+\theta_4)x}\right)$$

$$+ \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_2)(\theta_4-\theta_3)}\left(e^{-(\theta_1+\theta_2+\theta_3+\theta_4)x} - e^{-(\theta_1+\theta_2+2\theta_3)x}\right)$$

$$E(\hat{\lambda}_3) = \frac{1}{\theta_3} - \frac{\theta_3}{\theta_3-\theta_2}\left(\frac{1}{\theta_2+\theta_3} - \frac{1}{2\theta_2}\right) - \frac{\theta_2\theta_3}{(\theta_2-\theta_1)(\theta_3-\theta_1)}\left(\frac{1}{2\theta_1+\theta_3} - \frac{1}{3\theta_1}\right)$$

$$- \frac{\theta_2\theta_3}{(\theta_2-\theta_1)(\theta_3-\theta_2)}\left(\frac{1}{\theta_1+2\theta_2} - \frac{1}{\theta_1+\theta_2+\theta_3}\right)$$

$$E(\hat{\lambda}_4) = \frac{1}{\theta_4} - \frac{\theta_4}{\theta_4-\theta_3}\left(\frac{1}{\theta_3+\theta_4} - \frac{1}{2\theta_3}\right) - \frac{\theta_3\theta_4}{(\theta_3-\theta_2)(\theta_4-\theta_2)}\left(\frac{1}{2\theta_2+\theta_4} - \frac{1}{3\theta_2}\right)$$

$$- \frac{\theta_3\theta_4}{(\theta_3-\theta_2)(\theta_4-\theta_3)}\left(\frac{1}{\theta_2+2\theta_3} - \frac{1}{\theta_2+\theta_3+\theta_4}\right)$$

$$- \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_1)(\theta_4-\theta_1)}\left(\frac{1}{3\theta_1+\theta_4} - \frac{1}{4\theta_1}\right)$$

$$- \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_1)(\theta_4-\theta_3)}\left(\frac{1}{2(\theta_1+\theta_3)} - \frac{1}{2\theta_1+\theta_3+\theta_4}\right)$$

$$- \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_2)(\theta_4-\theta_2)}\left(\frac{1}{\theta_1+3\theta_2} - \frac{1}{\theta_1+2\theta_2+\theta_4}\right)$$

$$- \frac{\theta_2\theta_3\theta_4}{(\theta_2-\theta_1)(\theta_3-\theta_2)(\theta_4-\theta_3)}\left(\frac{1}{\theta_1+\theta_2+\theta_3+\theta_4} - \frac{1}{\theta_1+\theta_2+2\theta_3}\right).$$

APPENDIX B

This appendix contains proofs of the continuity of $a_k(\theta_1, \theta_2, \ldots, \theta_k)$ (Section 2.2) and of Lemma 1, Section 2.3.

Theorem B: Let $X_1, X_2, \ldots, X_k$ be independent exponentially distributed lifetimes with means $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$. Let

$$\hat{\lambda}_k = \max\left(X_k, \frac{X_k+X_{k-1}}{2}, \ldots, \bar{X}\right)$$

and

$$a_k(\theta_1, \theta_2, \ldots, \theta_k) = P\{\hat{\lambda}_k \leq x\}$$

where $\theta_i = 1/\lambda_i$ for $i = 1, 2, \ldots, k$ and x is considered fixed. Then

$$\lim_{\theta_i \downarrow \theta_{i+1}} a_k(\theta_1, \theta_2, \ldots, \theta_k) = a_k(\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \theta_{i+1}, \theta_{i+2}, \ldots, \theta_k) .$$

Proof: Let $Y_1, Y_2, \ldots, Y_k$ be iid exponential random variables with mean 1, and let $X_i = Y_i/\theta_i$, $i = 1, 2, \ldots, k$. Then X_i has an exponential distribution with mean $\lambda_i = 1/\theta_i$. Let $\theta_i^{(1)}, \theta_i^{(2)}, \ldots,$ be any nondecreasing sequence converging to θ_{i+1}, and define

$X_i^{(j)} = \dfrac{Y_i}{\theta_i^{(j)}}$, $j = 1, 2, 3, \ldots$. Let $\hat{\lambda}_k^{(j)}$ be the isotonic estimate of λ_k based on $X_1, X_2, \ldots, X_{i-1}, X_i^{(j)}, X_{i+1}, \ldots, X_k$.

Now define $X_i^{(\infty)} = \lim_{j\to\infty} X_i^{(j)} = \lim_{j\to\infty} \dfrac{y_i}{\theta_i^{(j)}} = \dfrac{y_i}{\theta_{i+1}}$ and let $\hat{\lambda}_k^{(\infty)}$ be the isotonic estimate of λ_k based on $X_1, X_2, \ldots, X_{i-1}, X_i^{(\infty)}, X_{i+1}, \ldots, X_k$. It is seen by the definition of $\hat{\lambda}_k$ that $\hat{\lambda}_k$ is a continuous nondecreasing function in each X_i, so $\hat{\lambda}_k^{(j)} \uparrow \hat{\lambda}_k^{(\infty)}$ since $X_i^{(j)} \uparrow X_i^{(\infty)}$. Now apply a monotone convergence theorem (Royden, page 85 (1968)) to obtain the result

$$\lim_{j\to\infty} P\{\hat{\lambda}_k^{(j)} \leq x\} = P\{\lim_{j\to\infty} \hat{\lambda}_k^{(j)} \leq x\}$$

$$= P\{\hat{\lambda}_k^{(\infty)} \leq x\} \ .$$

Equivalently, we have

$$\lim_{j\to\infty} a_k(\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_i^{(j)}, \theta_{i+1}, \ldots, \theta_k)$$

$$= a_k(\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \theta_{i+1}, \theta_{i+2}, \ldots, \theta_k) \ .$$

Since $\theta_i^{(j)}$ is an arbitrary sequence converging to θ_{i+1}, this concludes the proof. □

Lemma 1: Define:

$$c_k^{(r)} = \int_0^1 \int_{y_{k/2}}^1 \cdots \int_{[k-1/k]y_2}^1 y_1^r \, dy_1 \, dy_2 \cdots dy_k \, .$$

Then

$$c_k^{(r)} = \frac{1}{r+1}\left[\frac{k^{k-2}}{[(k-1)!]^2} + \sum_{i=1}^{k-1} (-1)^i \frac{(k-i)^{k+r}}{(k-i)!} \frac{(r+1)!}{(r+i+1)!} \frac{1}{k^r k!}\right] \tag{1}$$

where $k \geq 2$ and $r \geq 0$.

Proof: Evaluate the innermost integral of $c_k^{(r)}$ to obtain

$$c_k^{(r)} = \frac{1}{r+1} \int_0^1 \int_{y_{k/2}}^1 \cdots \int_{[(k-2)/(k-1)]y_3}^1 \left[1 - \left(\frac{k-1}{k}\right)^{r+1} y_2^{r+1}\right] dy_2 \cdots dy_k$$

$$= \frac{1}{r+1}\left[c_{k-1}^{(0)} - \left(\frac{k-1}{k}\right)^{r+1} c_{k-1}{}^{(r+1)}\right] \, . \tag{2}$$

This recursive relationship and the principle of mathematical induction (on k) will be used to prove the result. Consequently, we need to show that expression (1) holds for $k = 2$ and that it satisfies the relation given in expression (2).

When $k = 2$ we have, from the definition of $c_k^{(r)}$,

$$c_2^{(r)} = \int_0^1 \int_{y_{2/2}}^1 y_1^r \, dy_1 \, dy_2$$

$$= \frac{1}{r+1} \int_0^1 \left[1 - \left(\frac{1}{2}\right)^{r+1} y_2^{\,r+1} \right] dy_2$$

$$= \frac{1}{r+1} \left[1 - \left(\frac{1}{2}\right)^{r+1} \left(\frac{1}{r+2}\right) \right] .$$

It is easily seen that the same result is obtained by letting $k = 2$ in expression (1). It remains to show that (1) satisfies the relation given in (2). Substitution of (1) into the right-hand-side of (2) gives

$$\frac{1}{r+1} \left[\frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{i=1}^{k-2} (-1)^i \frac{(k-1-i)^{k-1}}{(k-1-i)!} \frac{1}{(i+1)!(k-1)!} \right] - \frac{1}{r+1} \left(\frac{k-1}{k}\right)^{r+1}$$

$$\frac{1}{r+2} \left[\frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{i=1}^{k-2} (-1)^i \frac{(k-1-i)^{k+r}}{(k-1-i)!} \frac{(r+2)!}{(r+i+2)!} \frac{1}{(k-1)^{r+1}(k-1)!} \right] . \quad (3)$$

Working with the first term in expression (3), we obtain

$$\frac{1}{r+1} \left[\frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{i=2}^{k-1} (-1)^{i-1} \frac{(k-i)^{k-1}}{(k-i)!} \frac{1}{i!(k-1)!} \right]$$

(raise summation index by 1)

$$= \frac{1}{r+1} \left[\sum_{i=1}^{k-1} (-1)^{i-1} \frac{(k-i)^{k-1}}{(k-i)!} \frac{1}{i!(k-1)!} \right]$$

(incorporate the first term into the sum)

$$= \frac{1}{r+1}\left[\sum_{j=1}^{k-1} (-1)^{k-j-1} \frac{j^{k-1}}{j!} \frac{1}{(k-j)!(k-1)!}\right]$$

(letting $j=k-i$)

$$= \frac{1}{r+1}\left[\sum_{j=0}^{k} (-1)^{k-j-1} \frac{j^{k-1}}{j!} \frac{1}{(k-j)!(k-1)!} + \frac{k^{k-1}}{k!(k-1)!}\right]$$

(add and subtract the $j=0$ and $j=k$ terms)

$$= \frac{1}{r+1}\left[\frac{(-1)}{(k-1)!k!} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} j^{k-1} + \frac{k^{k-2}}{[(k-1)!]^2}\right] .$$

The summation given above is, by a combinatorial identity (Feller, Vol. I, page 65 [8]), equal to zero. Therefore, the first term in expression (3) reduces to

$$\frac{1}{r+1} \frac{k^{k-2}}{[(k-1)!]^2} . \tag{4}$$

Raising the summation index by 1, we obtain for the second term in expression (3):

$$- \frac{1}{r+1}\left(\frac{k-1}{k}\right)^{r+1} \frac{1}{r+2}\left[\frac{(k-1)^{k-3}}{[(k-2)!]^2} + \sum_{i=2}^{k-1} (-1)^{i-1} \frac{(k-i)^{k+r}}{(k-i)!} \frac{(r+2)!}{(r+i+1)!} \frac{1}{(k-1)^{r+1}(k-1)!}\right]$$

$$= \frac{1}{r+1}\left[\frac{-(k-1)^{k+r-2}}{k^{r+1}(r+2)[(k-2)!]^2} - \sum_{i=2}^{k-1} (-1)^{i-1} \frac{(k-i)^{k+r}}{(k-i)!} \frac{(r+1)!}{(r+i+1)!} \frac{1}{k^r k!}\right]$$

$$= \frac{1}{r+1}\left[\frac{-(k-1)^{k+r}}{k^r(r+2)k!(k-1)!} + \sum_{i=2}^{k-1}(-1)^i \frac{(k-i)^{k+r}}{(k-i)!}\frac{(r+1)!}{(r+i+1)!}\frac{1}{k^r k!}\right]$$

$$= \frac{1}{r+1}\sum_{i=1}^{k-1}(-1)^i \frac{(k-i)^{k+r}}{(k-i)!}\frac{(r+1)!}{(r+i+1)!}\frac{1}{k^r k!} \qquad (5)$$

Now note that the sum of (4) and (5) is equal to (1). Therefore, we have shown that expression (1) holds for $k = 2$ and satisfies the recursive relation (2), so we conclude by the principle of induction that (1) holds for all values of k. □

BIBLIOGRAPHY

[1] Barlow, R.E., D.J. Bartholomew, J.M. Bremner, and H.D. Brunk, Statistical Inference Under Order Restrictions, (Wiley, 1972).

[2] Barlow, R.E., and F. Proschan, Statistical Theory of Reliability and Life Testing, (Holt, Rinehart, and Winston, (1975).

[3] Barlow, R.E., and E.M. Scheuer, "Reliability Growth During a Development Testing Program," Technometrics, vol. 8 (1966), pp. 53-60.

[4] Bresenham, J.E., "Reliability Growth Models," Technical Report #74 (Department of Statistics, Stanford University, 1964).

[5] Cohen, A., and H.B. Sackrowitz, "Estimation of the Last Mean of a Monotone Sequence," Annals of Mathematical Statistics, vol. 41 (1970), pp. 2021-2034.

[6] Corcoran, W.J., H. Weingarten, and P.W. Zehna, "Estimating Reliability After Corrective Action," Management Science, vol. 10 (1964), pp. 786-795.

[7] Davis, H.T., "Some Results on Isotonic Regression," Technical Report, (University of Rochester, 1977).

[8] Feller, W., An Introduction to Probability Theory and Its Applications, vol. I (Wiley, 1968).

[9] Feller, W., An Introduction to Probability Theory and Its Applications, vol. II (Wiley, 1971).

[10] Ferguson, T.S., Mathematical Statistics: A Decision Theoretic Approach, (Academic Press, 1967).

[11] Hogg, R.V., and E.G. Malmgren, "Practical Estimators of Ordered Parameters that Reduce Mean Square Error," Technical Report (University of Iowa, 1972).

[12] Jayachandran, T., and L.R. Moore, "Comparison of Reliability Growth Models," IEEE Transactions on Reliability, vol. 25 (1976), pp. 49-51.

[13] Langberg, N., and F. Proschan, "A Reliability Growth Model Involving Dependent Components," FSU Statistics Report M437 (Florida State University, 1977).

[14] Littlewood, B., and J. Verral, "A Bayesian Reliability Growth Model for Computer Software," Applied Statistics, vol. 22 (1973), pp. 332-346.

[15] Mood, A.M., F.A. Graybill, and D.C. Boes, Introduction to the Theory of Statistics, (McGraw-Hill, 1974).

[16] Parsons, L.V., "Distribution Theory of Isotonic Estimators," Ph.D. Thesis, (University of Iowa, 1975).

[17] Robertson, T., and F.T. Wright, "On Estimating Monotone Parameters," Annals of Mathematical Statistics, vol. 39 (1968), pp. 1030-1039.

[18] Smith, A.F.M., "Bayesian Note on Reliability Growth During a Development Testing Program," IEEE Transactions on Reliability, vol. 26 (1977), pp. 346-347.

[19] Sposito, V.A., Linear and Nonlinear Programming, (Iowa University Press, 1975).

[20] Weinrich, M.C., and A.J. Gross, "Barlow-Scheuer Reliability Growth Model from a Bayesian Viewpoint," Technometrics, vol. 20 (1978), pp. 249-254.

TABLE 1

r	$E(\bar{X})$	$E(\lambda_k)$	$Var(\bar{X})$	$Var(\lambda_k)$	$MSE(\bar{X})$	$MSE(\lambda_k)$
1.00	1.000	1.464	.200	.785	.200	1.000
1.05	.909	1.402	.166	.793	.174	.954
1.10	.834	1.352	.142	.803	.169	.927
1.25	.672	1.252	.099	.833	.207	.897
1.50	.521	1.166	.071	.869	.300	.897
2.00	.388	1.093	.053	.909	.428	.918
4.00	.266	1.025	.043	.965	.581	.965
6.00	.240	1.012	.041	.981	.619	.981
8.00	.228	1.007	.041	.988	.636	.988
10.00	.222	1.004	.040	.992	.645	.992
100.00	.202	1.000	.040	.999	.677	.999

Number of stages = k = 5 $\qquad \lambda_k = 1.00 \qquad r = \frac{\lambda_{i+1}}{\lambda_i}$

TABLE 2

Percentage points for the distribution of $\hat{\lambda}_k$ in the equal-mean case.

α and (θx) satisfy the relationship $P_{\underline{\lambda}}\{\hat{\lambda}_k > \theta x\} = \alpha$, where $\lambda_1 = \cdots = \lambda_k$.

α	k=2	k=3	k=4	k=5	k=6
.25	1.66	1.75	1.79	1.81	1.83
.20	1.86	1.94	1.97	1.99	2.00
.15	2.12	2.19	2.21	2.22	2.23
.10	2.49	2.53	2.55	2.55	2.56
.05	3.12	3.15	3.15	3.15	3.15
.01	4.65	4.65	4.65	4.65	4.65

α	k=7	k=8	k=10	$k \geq 20$
.25	1.83	1.84	1.84	1.85
.20	2.00	2.01	2.01	2.01
.15	2.23	2.23	2.23	2.23
.10	2.56	2.56	2.56	2.56
.05	3.15	3.15	3.15	3.15
.01	4.65	4.65	4.65	4.65

TABLE 3

Concave Isotonic Estimates (k=3)

I. $x_1 \le x_2 \le x_3,\ x_3 - x_2 \le x_2 - x_1$ (Increasing, concave)

$$\lambda_1^* = x_1 \quad \lambda_2^* = x_2 \quad \lambda_3^* = x_3$$

II. $x_1 \le x_2 \le x_3,\ x_3 - x_2 > x_2 - x_1$ (Increasing, not concave)

$$\lambda_1^* = \frac{5x_1+2x_2-x_3}{6} \quad \lambda_2^* = \bar{x} \quad \lambda_3^* = \frac{-x_1+2x_2+5x_3}{6}$$

III(a). $x_1 \le x_2,\ x_2 > x_3,\ x_1 \le \frac{x_2+x_3}{2}$ (Increasing, decreasing)

$$\lambda_1^* = x_1,\ \lambda_2^* = \lambda_3^* = \frac{x_2+x_3}{2}$$

III(b). $x_1 \le x_2,\ x_2 > x_3,\ x_1 > \frac{x_2+x_3}{2}$

$$\lambda_1^* = \lambda_2^* = \lambda_3^* = \bar{x}$$

IV. $x_3 < x_2 < x_1$ (Decreasing)

$$\lambda_1^* = \lambda_2^* = \lambda_3^* = \bar{x}$$

V(a). $x_3 \ge x_1 > x_2$ (Decreasing, increasing)

Same solutions as in Case II.

V(b). $x_1 \ge x_3 \ge x_2$

$$\lambda_1^* = \lambda_2^* = \lambda_3^* = \bar{x}$$

TABLE 4

MSE Breakdown into Bias and Variance Components when Estimating Reliability at Various Times

$k = 5$ 1 sample per stage $\lambda_i = 10 - 8/i + 8/5$

$\hat{\lambda}_5$ = isotonic estimate of λ_5 $\bar{X}$ = overall mean estimate of λ_5

$E(\hat{\lambda}_5) = 13.71$ $Var(\hat{\lambda}_5) = 62.77$ $MSE(\hat{\lambda}_5) = 76.52$

$E(\bar{X}) = 7.95$ $Var(\bar{X}) = 13.71$ $MSE(\bar{X}) = 17.92$

		$e^{-t/\hat{\lambda}_k}$				$e^{-t/\bar{X}}$			
Time	True Reliability	Mean	$Bias^2$	Variance	MSE	Mean	$Bias^2$	Variance	MSE
3	.7408	.7473	.00004	.0178	.0178	.6511	.0081	.0182	.0262
6	.5488	.5762	.00075	.0318	.0326	.4421	.0114	.0244	.0358
9	.4066	.4540	.0022	.0361	.0383	.3086	.0096	.0214	.0310
12	.3012	.3637	.0039	.0350	.0389	.2198	.0066	.0164	.0231
15	.2231	.2952	.0052	.0316	.0368	.1591	.0041	.0119	.0160
18	.1653	.2421	.0059	.0275	.0334	.1166	.0024	.0084	.0107
21	.1224	.2004	.0061	.0234	.0295	.0864	.0013	.0058	.0071
24	.0907	.1672	.0058	.0197	.0255	.0647	.0007	.0040	.0047
27	.0672	.1405	.0054	.0164	.0218	.0488	.0003	.0028	.0031
30	.0498	.1187	.0047	.0136	.0184	.0372	.0002	.0019	.0021

TABLE 5

Relating Variance to Shape Parameter β for Weibull Distribution with Fixed Means

$k = 5$ stages

	Means at each stage				
β	3.60	7.60	8.93	9.60	10.00
1.00	12.96	57.76	79.74	92.16	100.00
1.25	8.40	37.43	51.72	59.72	64.80
1.50	5.98	26.63	36.79	42.49	46.10
1.75	4.51	20.09	27.76	32.06	34.78
2.00	3.54	15.78	21.81	25.18	27.32
2.25	2.87	12.77	17.65	20.38	22.11
2.50	2.37	10.58	14.61	16.88	18.31
2.75	2.00	8.92	12.32	14.23	15.44
3.00	1.71	7.63	10.54	12.17	13.20

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 194	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROPERTIES OF ISOTONIC ESTIMATORS OF MEAN LIFETIME IN A SIMPLE PROTOTYPE DEVELOPMENT MODEL		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas P. McWilliams		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0561
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research and Department of Statistics - Stanford University, Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-047-200)
11. CONTROLLING OFFICE NAME AND ADDRESS Operations Research, Code 434 Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE August 16, 1979 13. NUMBER OF PAGES 105
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Reliability growth
Isotonic regression
Prototype development
Mean lifetime estimation
Concave isotonic regression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consideration of a simple prototype development/reliability growth model leads to observation of variables $X_1^*, X_2^*, \ldots, X_k^*$; where the X_i^* are independent and are assumed to have the same distribution type with individual means λ_i. The problem is to estimate the λ_i, which are assumed to be non-decreasing due to design improvements in the device under development. For → next page

lambda sub i

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Technical Report No. 194, Abstract continued.

many common distributions, estimation via the restricted maximum likelihood principle leads to estimates which are the isotonic regression of the X_i^n with appropriate weights. This paper finds the distribution of $\hat{\lambda}_k$, the mle of λ_k, when the X_i^n are exponentially distributed. $\hat{\lambda}_k$ is then compared with $\bar{X}$ and X_k^n, two competing estimates of λ_k. Gamma, normal, and Weibull distributed X_i^n are also considered, and the usefulness of $\hat{\lambda}_k$ is seen to depend on the relative magnitude of the variances and means of the X_i^n. Results are also examined when the restricted mles are generated assuming the λ_i are a non-decreasing and concave function of i.

lambda sub i

lambda cap sub k

lambda sub k

X-bar

#194

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)